Access DB# 68180

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # : _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 – 703-308-4498
jan.delaval@uspto.gov

****

STAFF USE ONLY

Searcher: Jan
Searcher Phone #: 4498
Searcher Location: _____
Date Searcher Picked Up: 6/6/02
Date Completed: 6/6/02
Searcher Prep & Review Time: _____
Clerical Prep Time: 15
Online Time: 10

Type of Search

NA Sequence (#) 3
AA Sequence (#) 4
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems ✓
WWW/Internet _____
Other (specify) _____

PTO-1590 (8-01)

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:37 ; Search time 2149.71 Seconds
(without alignments)
13054.080 Million cell updates/sec

Title: US-08-881-509-1
Perfect score: 1341
Sequence: 1 ATGAGGCAAGTGGCGAGAGT.....GCCGATGCCTTCATTAAAAT 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1341	100.0	1341	6	A93127	A93127 Sequence 1
2	972.2	72.5	1508	9	HSTCRAR	X02592 Human mRNA
3	972.2	72.5	1524	9	HUMTCAXB	M12959 Human T-cel
4	958.2	71.5	1255	9	BC020840	BC020840 Homo sapi
5	950.8	70.9	1457	9	S72942	S72942 c6.1A...TCR
6	950.8	70.9	1647	9	S72931	S72931 Homo sapien
7	835.4	62.3	874	9	HUMATCRA	M13052 Human T-cel
8	691.4	51.6	1367	9	AK026255	AK026255 Homo sapi
9	683.2	50.9	1131	9	HSTCELA1	X01403 H.sapiens m
10	596.4	44.5	1064	9	HSTCARA	X63455 H.sapiens m
11	569.6	42.5	1080	9	HUMTCAYE	M15565 Human T-cel
12	538.8	40.2	840	9	HSTCRAC4	X05002 Human DNA f
13	538.8	40.2	840	9	HUMTCACA4	M14861 Human T-cel
14	535.6	39.9	208953	9	CNS01RH3	AL160314 Human chr
15	534	39.8	175053	2	AC023226	AC023226 Homo sapi
16	514.8	38.4	1081	9	HUMTCAXE	M12423 Human T-cel
17	514.8	38.4	1101	6	E00990	E00990 cDNA encodi
18	510	38.0	71153	9	HUAE000662	AE000662 Homo sapi
19	510	38.0	97630	9	HUMTCRADCV	M94081 Human Tcr-C
20	498.2	37.2	5177	9	HSTCRAC	X02883 Human gene
21	448	33.4	822	6	I14073	I14073 Sequence 8
22	446.4	33.3	834	6	AX201131	AX201131 Sequence
23	446.4	33.3	834	6	AX267930	AX267930 Sequence
24	444.6	33.2	804	9	AF043177	AF043177 Homo sapi
25	441.6	32.9	828	9	HSJ004871	AJ004871 Homo sapi
26	440.4	32.8	810	9	HSTCRDF8	Z49903 H.sapiens m
27	440.2	32.8	813	9	AF043174	AF043174 Homo sapi
28	439.8	32.8	828	6	AX322052	AX322052 Sequence
29	439.2	32.8	840	9	HUMTCRAZ	L34703 Homo sapien
30	438.2	32.7	858	9	AF043171	AF043171 Homo sapi
31	428	31.9	728	9	HUMTCAXA	K02777 Human T-cel
32	423	31.5	828	12	XXU40407	U40407 Synthetic c
33	420.2	31.3	426	9	HUMTCRA29Z	L02424 Homo sapien
34	393.4	29.3	400	11	G13536	G13536 SHGC-11082
35	382.4	28.5	1410	4	SHPTCRA	M55622 Sheep T cel
36	371.6	27.7	989	4	BOVTCRA21	D10394 Bovine T-ce
37	335	25.0	397	9	HUMTCAYX	M17663 Human T-cel
38	307.8	23.0	744	6	AX010058	AX010058 Sequence
39	307.8	23.0	744	6	AX010083	AX010083 Sequence
40	307.8	23.0	744	6	AX010104	AX010104 Sequence
41	307.8	23.0	744	6	AX010157	AX010157 Sequence
42	307.2	22.9	765	6	AX010074	AX010074 Sequence
43	307.2	22.9	765	6	AX010093	AX010093 Sequence
44	307.2	22.9	765	6	AX010114	AX010114 Sequence
45	307.2	22.9	765	6	AX010173	AX010173 Sequence

ALIGNMENTS

```
RESULT   1
A93127
LOCUS       A93127                  1341 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0816496.
ACCESSION   A93127
VERSION     A93127.1  GI:6741516
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1341)
  AUTHORS   Schendel,D.J.
  TITLE     T-cells specific for kidney carcinoma
  JOURNAL   Patent: EP 0816496-A 1 07-JAN-1998;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .1341
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     sig_peptide     1. .54
     CDS             1. .804
                     /note="unnamed protein product"
                     /codon_start=1
```

```
                     /protein_id="CAB69529.1"
                     /db_xref="GI:6741517"
                     /translation="MRQVARVIVFLTLSTLSLAKTTQPISMDSYEGQEVNITCSHNNI
                     ATNDYITWYQQFPSQGPRFIIQGYKTKVTNEVASLFIPADRKSSTLSLPRVSLSDTAV
                     YYCLVGGSARQLTFGSGTQLTVLPDIQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNV
                     SQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPE
                     SSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS"
     mat_peptide     55. .801
                     /product="unnamed"
BASE COUNT      331 a    365 c    294 g    351 t
ORIGIN
```

Query Match 100.0%; Score 1341; DB 6; Length 1341;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCTGAGTACTTTGAGCCTTGCTAAG 60
Db       1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCTGAGTACTTTGAGCCTTGCTAAG 60

Qy      61 ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGC 120
Db      61 ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGC 120

Qy     121 CACAACAACATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA 180
Db     121 CACAACAACATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA 180

Qy     181 CCACGATTTATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTT 240
Db     181 CCACGATTTATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTT 240

Qy     241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACT 300
Db     241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACT 300

Qy     301 GCTGTGTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360
Db     301 GCTGTGTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360

Qy     361 CAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGAC 420
Db     361 CAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGAC 420

Qy     421 TCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTG 480
Db     421 TCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTG 480

Qy     481 TCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCT 540
Db     481 TCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCT 540

Qy     541 ATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCA 600
Db     541 ATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCA 600

Qy     601 AACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
Db     601 AACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660

Qy     661 TGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC 720
Db     661 TGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC 720

Qy     721 CTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATG 780
Db     721 CTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATG 780

Qy     781 ACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCG 840
Db     781 ACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCG 840

Qy     841 CTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCC 900
Db     841 CTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCC 900

Qy     901 AAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCT 960
Db     901 AAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCT 960

Qy     961 ACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGT 1020
Db     961 ACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGT 1020

Qy    1021 TCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCT 1080
Db    1021 TCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCT 1080

Qy    1081 CCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATG 1140
Db    1081 CCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATG 1140

Qy    1141 ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTAT 1200
Db    1141 ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTAT 1200

Qy    1201 TTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGG 1260
Db    1201 TTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGG 1260

Qy    1261 AAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGC 1320
Db    1261 AAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGC 1320

Qy    1321 TGCCGATGCCTTCATTAAAAT 1341
Db    1321 TGCCGATGCCTTCATTAAAAT 1341
```

```
RESULT   2
HSTCRAR
LOCUS       HSTCRAR                 1508 bp    mRNA    linear   PRI 29-JAN-1995
DEFINITION  Human mRNA for T-cell receptor alpha chain (TCR-alpha).
ACCESSION   X02592 M12959
VERSION     X02592.1  GI:36944
KEYWORDS    membrane protein; T-cell receptor; T-cell receptor alpha.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1508)
  AUTHORS   Rabbitts,T.H., Lefranc,M.P., Stinson,M.A., Sims,J.E., Schroder,J.,
            Steinmetz,M., Spurr,N.L., Solomon,E. and Goodfellow,P.N.
  TITLE     The chromosomal location of T-cell receptor genes and a T cell
            rearranging gene: possible correlation with specific translocations
            in human T cell leukaemia
  JOURNAL   EMBO J. 4 (6), 1461-1465 (1985)
  MEDLINE   85284934
FEATURES             Location/Qualifiers
     source          1. .1508
                     /organism="Homo sapiens"
                     /strain="T-cell leukaemic cell line JM"
                     /db_xref="taxon:9606"
     CDS             129. .962
                     /codon_start=1
                     /product="TCR-alpha chain"
                     /protein_id="CAA26435.1"
                     /db_xref="GI:36945"
                     /translation="MLLLLVPVLEVIFTLGGTRAQSVTQLGSHVSVSEGALVLLRCNY
                     SSSVPPYLFWYVQYPNQGLQLLLKYTSAATLVKGINGFEAEFKKSETSFHLTKPSAHM
                     SDAAEYFCAVSDLEPNSSASKIIFGSGTRLSIRPNIQNPDPAVYQLRDSKSSDKSVCL
                     FTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSII
                     PEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWS
                     S"
     misc_feature    468. .485
```

```
     misc_feature    /note="pot. variable region (aa 114-119)"
                     486. .533
     misc_feature    /note="pot. joining region (aa 120-135)"
                     534. .959
     misc_feature    /note="pot. constant region (aa 136-277)"
                     1492. .1497
                     /note="pot. polyA signal"
BASE COUNT      344 a    431 c    341 g    392 t
ORIGIN
```

Query Match 72.5%; Score 972.2; DB 9; Length 1508;
Best Local Similarity 98.2%; Pred. No. 2.3e-300;
Matches 983; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy    341 TGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTG 400
          | | |||||||||| |||||   | | |   |    ||  ||||||||||||||||||||
Db    499 TAATCTTTGGATCAGGGACCAGACTCAGCATCCGGCCAAATATCCAGAACCCTGACCCTG 558

Qy    401 CCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATT 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 CCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATT 618

Qy    461 TTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAA 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 TTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAA 678

Qy    521 CTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACA 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 CTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACA 738

Qy    581 AATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCT 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 AATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCT 798

Qy    641 TCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATA 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 TCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATA 858

Qy    701 CGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGG 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    859 CGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGG 918

Qy    761 CCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTA 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    919 CCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTA 978

Qy    821 AGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACA 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    979 AGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACA 1038

Qy    881 GAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGC 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1039 GAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGT 1098

Qy    941 AATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAAC 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1099 AATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAAC 1158

Qy   1001 ACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAG 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1159 ACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAG 1218

Qy   1061 AGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTG 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1219 AGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTG 1278

Qy   1121 CCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGA 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1279 CCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGA 1338

Qy   1181 GAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCT 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1339 GAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCT 1398

Qy   1241 TCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCT 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1399 TCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCT 1458

Qy   1301 GGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1341
          |||||||||||||||||||||||||||||||||||||||||
Db   1459 GGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1499
```

```
RESULT   3
HUMTCAXB
LOCUS       HUMTCAXB                1524 bp    mRNA    linear   PRI 13-JAN-1995
DEFINITION  Human T-cell receptor active alpha-chain mRNA from JM cell line,
            complete cds.
ACCESSION   M12959
VERSION     M12959.1  GI:338734
KEYWORDS    C-region; D-region; J-region; T-cell receptor; V-region; antigen
            receptor.
SOURCE      Human T-cell leukaemic cell line JM, cDNA to mRNA, clone pJM3E11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Rabbitts,T.H., Lefranc,M.P., Stinson,M.A., Sims,J.E., Schroder,J.,
            Steinmetz,M., Spurr,N.L., Solomon,E. and Goodfellow,P.N.
  TITLE     The chromosomal location of T-cell receptor genes and a T cell
            rearranging gene: possible correlation with specific translocations
            in human T cell leukaemia
  JOURNAL   EMBO J. 4 (6), 1461-1465 (1985)
  MEDLINE   85284934
COMMENT     A clean copy of the sequence in [1] was kindly provided by
            M.P.Lefranc, 01-AUG-1986.
            A polyadenylation signal was identified at positions 1500-1506.
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
     gene            1. .1524
                     /gene="TCRA"
     CDS             137. .970
                     /gene="TCRA"
                     /note="T-cell receptor alpha-chain (VDJC)"
                     /codon_start=1
                     /protein_id="AAA60625.1"
                     /db_xref="GI:338735"
                     /db_xref="GDB:G00-120-404"
                     /translation="MLLLLVPVLEVIFTLGGTRAQSVTQLGSHVSVSERALVLLRCNY
                     SSSVPPYLFWYVQYPNQGLQLLLKYTSAATLVKGINGFEAEFKKSETSFHLTKPSAHM
                     SDAAEYFCAVSDLEPNSSASKIIFGSGTRLSIRPNIQNPDPAVYQLRDSKSSDKSVCL
                     FTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSII
                     PEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWS
                     S"
     misc_recomb     475. .476
                     /gene="TCRA"
                     /organism="Homo sapiens"
     misc_recomb     493. .494
                     /gene="TCRA"
                     /organism="Homo sapiens"
     misc_recomb     541. .542
                     /gene="TCRA"
                     /organism="Homo sapiens"
BASE COUNT      347 a    437 c    345 g    395 t
ORIGIN      341 bp upstream of RsaI site; chromosome 14.
```

Query Match 72.5%; Score 972.2; DB 9; Length 1524;
Best Local Similarity 98.2%; Pred. No. 2.3e-300;
Matches 983; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy    341 TGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTG 400
```

```
Db    507 TAATCTTTGGATCAGGGACCAGACTCAGCATCCGGCCAAATATCCAGAACCCTGACCCTG 566
Qy    401 CCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATT 460
Db    567 CCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATT 626
Qy    461 TTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAA 520
Db    627 TTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAA 686
Qy    521 CTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACA 580
Db    687 CTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACA 746
Qy    581 AATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCT 640
Db    747 AATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCT 806
Qy    641 TCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATA 700
Db    807 TCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATA 866
Qy    701 CGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGG 760
Db    867 CGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGG 926
Qy    761 CCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTA 820
Db    927 CCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTA 986
Qy    821 AGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACA 880
Db    987 AGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACA 1046
Qy    881 GAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGC 940
Db   1047 GAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGT 1106
Qy    941 AATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAAC 1000
Db   1107 AATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAAC 1166
Qy   1001 ACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAG 1060
Db   1167 ACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAG 1226
Qy   1061 AGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTG 1120
Db   1227 AGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTG 1286
Qy   1121 CCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGA 1180
Db   1287 CCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGA 1346
Qy   1181 GAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCT 1240
Db   1347 GAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCT 1406
Qy   1241 TCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCT 1300
Db   1407 TCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCT 1466
Qy   1301 GGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1341
Db   1467 GGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1507
```

```
RESULT   4
BC020840
LOCUS       BC020840                1255 bp    mRNA    linear   PRI 08-JAN-2002
DEFINITION  Homo sapiens, clone MGC:23964 IMAGE:4687209, mRNA, complete cds.
ACCESSION   BC020840
VERSION     BC020840.1  GI:18088582
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1255)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 36 Row: k Column: 12
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .1255
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:23964 IMAGE:4687209"
                     /tissue_type="Testis, embryonal carcinoma"
                     /clone_lib="NIH_MGC_61"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             66. .674
                     /codon_start=1
                     /product="Unknown (protein for MGC:23964)"
                     /protein_id="AAH20840.1"
                     /db_xref="GI:18088583"
                     /translation="MQRISSLIHLSLFWAGVMSAIELVPEHQTVPVSIGVPATLRCSM
                     KGEAIGNYYINWYRKTQDIQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSD
                     VYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPESSCDVKL
                     VEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS"
BASE COUNT      322 a    332 c    285 g    316 t    .
ORIGIN
```

```
Query Match           71.5%;  Score 958.2;  DB 9;  Length 1255;
Best Local Similarity  99.7%;  Pred. No. 7e-296;
Matches  960;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    379 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 438
Db    249 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 308
Qy    439 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 498
Db    309 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 368
Qy    499 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 558
Db    369 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 428
Qy    559 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 618
Db    429 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 488
```

```
Qy    619 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 548

Qy    679 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 608

Qy    739 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 668

Qy    799 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 728

Qy    859 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCT 918
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    729 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCCACCCCCAAGGAGGTGAAAGCTGCT 788

Qy    919 ACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAA 978
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 ACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAA 848

Qy    979 GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGT 1038
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGT 908

Qy   1039 CACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATT 1098
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 CACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATT 968

Qy   1099 CCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTT 1158
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 CCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTT 1028

Qy   1159 CTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTC 1218
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTC 1088

Qy   1219 ATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGA 1278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 ATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGA 1148

Qy   1279 GGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAA 1338
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1149 GGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAA 1208

Qy   1339 AAT 1341
          |||
Db   1209 AAT 1211
```

```
RESULT   5
S72942
LOCUS       S72942                  1457 bp    DNA     linear   PRI 26-JAN-1995
DEFINITION  c6.1A. . .TCR C alpha  {clone 2b, translocation breakpoint} [human,
            ataxia telangiectasia patient AT8BI, Genomic Mutant, 3 genes, 1457
            nt].
ACCESSION   S72942
VERSION     S72942.1  GI:639614
KEYWORDS    .
SOURCE      human ataxia telangiectasia patient AT8BI.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1457)
  AUTHORS   Thick,J., Mak,Y.F., Metcalfe,J., Beatty,D. and Taylor,A.M.
  TITLE     A gene on chromosome Xq28 associated with T-cell prolymphocytic
            leukemia in two patients with ataxia telangiectasia
  JOURNAL   Leukemia 8 (4), 564-573 (1994)
  MEDLINE   94202904
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 154727] from the original journal article.
            This sequence comes from Fig. 1.
            Map location: t(X;14)(q28;q11).
FEATURES             Location/Qualifiers
     source          1. .1457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .478
                     /partial
                     /gene="c6.1A"
     gene            1. .1457
                     /partial
                     /gene="c6.1A-TCR C&agr;"
     gene            479. .1457
                     /partial
                     /gene="TCR C&agr;"
BASE COUNT      390 a    364 c    339 g    364 t
ORIGIN
```

Query Match 70.9%; Score 950.8; DB 9; Length 1457;
Best Local Similarity 99.7%; Pred. No. 1.7e-293;
Matches 963; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy    377 CTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA 436
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 CAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA 535

Qy    437 AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATT 496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    536 AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATT 595

Qy    497 CTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    596 CTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 655

Qy    557 ACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACA 616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    656 ACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACA 715

Qy    617 GCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGG 676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    716 GCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGG 775

Qy    677 TCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGT 736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    776 TCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGT 835

Qy    737 TCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGT 796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    836 TCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGT 895

Qy    797 CCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATT 856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    896 CCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATT 955

Qy    857 GCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG 916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    956 GCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG 1015

Qy    917 CTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1016 CTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 1075

Qy    977 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTT 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1076 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTT 1135

Qy   1037 GTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACA 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1136 GTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACA 1195
```

```
Qy   1097 TTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGG 1156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1196 TTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGG 1255

Qy   1157 TTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGT 1216
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1256 TTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGT 1315

Qy   1217 TCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATC 1276
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1316 TCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATC 1375

Qy   1277 GAGGCCCTGCTATGCTGTGTG-TCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCAT 1335
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1376 GAGGCCCTGCTATGCTGTGTGATCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCAT 1435

Qy   1336 TAAAAT 1341
          ||||||
Db   1436 TAAAAT 1441
```

RESULT 6
S72931
LOCUS S72931 1647 bp DNA linear PRI 08-MAY-2000
DEFINITION Homo sapiens T-cell receptor alpha chain-c6.1A fusion protein (c6.1A-TCRC) gene, partial cds.
ACCESSION S72931
VERSION S72931.1 GI:639612
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1647)
AUTHORS Thick,J., Mak,Y.F., Metcalfe,J., Beatty,D. and Taylor,A.M.
TITLE A gene on chromosome Xq28 associated with T-cell prolymphocytic leukemia in two patients with ataxia telangiectasia
JOURNAL Leukemia 8 (4), 564-573 (1994)
MEDLINE 94202904
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 154715] from the original journal article.
This sequence comes from Fig. 1.
Map location: t(X;14)(q28;q11).

FEATURES Location/Qualifiers
source 1. .1647
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2a"
/note="ataxia telangiectasia patient AT8BI"
gene <1. .1093
/gene="c6.1A-TCRC"
CDS <1. .1093
/gene="c6.1A-TCRC"
/note="This sequence comes from Fig. 1; author's translation differs from conceptual translation"
/codon_start=2
/product="T-cell receptor alpha chain-c6.1A fusion protein"
/protein_id="AAB30469.2"
/db_xref="GI:7717235"
/translation="DTRSDSKFAYTGTEMRTVAEKVDAVRIVHIHSVIILRRSDKRKD
RVEISPEQLSAASTEAERLAELTGRPMRVVGWYHSHPHITVWPSHVDVRTQAMYQMMD
QGFVGLIFSCFIEDKNTKTGRVLYTCFQSIQAQKSSESLHGPRDFWSSSQHISIEGQK
EEERYERIEIPIHIVPHVTIGKVCLESAVELPKILCQEEQDRYRRIHSLTHLDSVTKI
HNGSDIQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRS
MDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPESSCDVKLVEKSFETDTNLNF
QNLSVIGFRILLLKVAGFNLLMTLRLWSS"
misc_recomb 669^670
/gene="c6.1A-TCRC"
/note="translocation breakpoint"

BASE COUNT 427 a 422 c 373 g 425 t
ORIGIN

Query Match 70.9%; Score 950.8; DB 9; Length 1647;
Best Local Similarity 99.7%; Pred. No. 1.8e-293;
Matches 963; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy    377 CTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA 436
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 CAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA 725

Qy    437 AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATT 496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    726 AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATT 785

Qy    497 CTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    786 CTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 845

Qy    557 ACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACA 616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    846 ACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACA 905

Qy    617 GCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGG 676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    906 GCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGG 965

Qy    677 TCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGT 736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    966 TCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGT 1025

Qy    737 TCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGT 796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 TCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGT 1085

Qy    797 CCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATT 856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1086 CCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATT 1145

Qy    857 GCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG 916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1146 GCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG 1205

Qy    917 CTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1206 CTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 1265

Qy    977 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTT 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1266 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTT 1325

Qy   1037 GTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACA 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1326 GTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACA 1385

Qy   1097 TTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGG 1156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1386 TTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGG 1445

Qy   1157 TTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGT 1216
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1446 TTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGT 1505

Qy   1217 TCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATC 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1506 TCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATC 1565

Qy   1277 GAGGCCCTGCTATGCTGTGTG-TCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCAT 1335
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1566 GAGGCCCTGCTATGCTGTGTGATCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCAT 1625

Qy   1336 TAAAAT 1341
          ||||||
Db   1626 TAAAAT 1631
```

```
RESULT   7
HUMATCRA
LOCUS       HUMATCRA                  874 bp    DNA     linear   PRI 31-OCT-1994
DEFINITION  Human T-cell antigen receptor alpha-chain mRNA, 3' end.
ACCESSION   M13052
VERSION     M13052.1  GI:179135
KEYWORDS    T cell receptor.
SOURCE      Human leukemic T-cell line Jurkat, cDNA to mRNA, clone pJ6-alpha-2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   Collins,M.K., Tanigawa,G., Kissonerghis,A.M., Ritter,M.,
            Price,K.M., Tonegawa,S. and Owen,M.J.
  TITLE     Regulation of T-cell receptor gene expression in human T-cell
            development
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 82 (13), 4503-4507 (1985)
  MEDLINE   85242713
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
     mRNA            <1. .>874
                     /product="T-cell antigen receptor mRNA"
     gene            1. .323
                     /gene="TCRA"
     CDS             <1. .323
                     /gene="TCRA"
                     /note="T-cell antigen receptor alpha-chain"
                     /codon_start=3
                     /protein_id="AAA51791.1"
                     /db_xref="GI:179136"
                     /db_xref="GDB:G00-120-404"
                     /translation="QSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNN
                     SIIPEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLR
                     LWSS"
BASE COUNT      206 a    238 c    195 g    235 t
ORIGIN      203 bp upstream of HindIII site; chromosome 14q11.2.
```

```
  Query Match           62.3%;  Score 835.4;  DB 9;  Length 874;
  Best Local Similarity  99.7%;  Pred. No. 1.8e-256;
  Matches  858;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy     482 CACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTA 541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTA 60

Qy     542 TGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAA 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAA 120

Qy     602 ACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCT 661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCT 180

Qy     662 GTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACC 721
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACC 240

Qy     722 TGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGA 781
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGA 300

Qy     782 CGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGC 841
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGC 360

Qy     842 TCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCA 901
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCA 420

Qy     902 AGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTA 961
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTA 480

Qy     962 CCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTT 1021
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTT 540

Qy    1022 CCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTC 1081
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTC 600

Qy    1082 CCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCAT-CCCACAGATG 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db     601 CCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCCACGGATG 660

Qy    1141 ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTAT 1200
         |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     661 ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCT-GAGAATGTTGTGAGGGGTTTAT 719

Qy    1201 TTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGG 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 TTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGG 779

Qy    1261 AAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGC 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 AAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGC 839

Qy    1321 TGCCGATGCCTTCATTAAAAT 1341
         |||||||||||||||||||||
Db     840 TGCCGATGCCTTCATTAAAAT 860
```

```
RESULT   8
AK026255
LOCUS       AK026255                  1367 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ22602 fis, clone HSI04504, highly similar to
            HSTCARA Homo sapiens mRNA for T-cell antigen receptor alpha-chain.
ACCESSION   AK026255
VERSION     AK026255.1  GI:10439050
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens human small intestine cDNA to mRNA, clone_lib:HSI
            clone:HSI04504.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1367)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="HSI04504"
                     /clone_lib="HSI"
                     /tissue_type="human small intestine"
                     /note="cloning vector pME18SFL3"
     misc_feature    1. .1367
                     /note="highly similar to HSTCARA Homo sapiens mRNA for
                     T-cell antigen receptor alpha-chain"
BASE COUNT      386 a    331 c    293 g    357 t
ORIGIN

  Query Match           51.6%;  Score 691.4;  DB 9;  Length 1367;
  Best Local Similarity 74.5%;  Pred. No. 3.4e-210;
  Matches 1004;  Conservative    0;  Mismatches  126;  Indels  217;  Gaps    2;

Qy      1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCTGAGTACTTTGAGCCTTGCTAAG 60
          ||||||| ||||| ||||| | ||||| |||||| | | ||| | |    |||||||
Db    198 ATGAGGCTGGTGGCAAGAGTAACTGTGTTTCTGACCTTTGGAACTATAATTGATGCTAAG 257

Qy     61 ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGC 120
          |||||||||||| |||||||| |    ||||||| || |  ||| |  | ||||
Db    258 ACCACCCAGCCCCCCTCCATGGATTGCGCTGAAGGAAGAGCTGCAAACCTGCCTTGTAAT 317

Qy    121 CACAACAACATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA 180
          |||   | |||   |  |||||| ||| |    ||||| | |||| ||| |  ||| ||
Db    318 CACTCTACCATCAGTGGAAATGAGTATGTGTATTGGTATCGACAGATTCACTCCCAGGGG 377

Qy    181 CCACGATTTATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTT 240
          ||||  | ||| ||||| ||    || |  || |  || || ||| ||||||| ||| |
Db    378 CCACAGTATATCATTCATGGTCTAAAAAACAATGAAACCAATGAAATGGCCTCTCTGATC 437

Qy    241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACT 300
          ||| | |  |||||||||||||||||  ||| ||||||||  | | | ||||| ||||||
Db    438 ATCACAGAAGACAGAAAGTCCAGCACCTTGATCCTGCCCCACGCTACGCTGAGAGACACT 497

Qy    301 GCTGTGTACTACTGCCTCGTGGG------TGGTTCTGCAAGGCAACTGACCTTTGGATCT 354
          |||||||||||| ||| ||||  |       | | | |    || |||||||||
Db    498 GCTGTGTACTATTGCATCGTCAGAGATAATAATGCAGGCAACATGCTCACCTTTGGAGGG 557

Qy    355 GGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTG 414
          || |||   || |  ||   ||| ||||||||||||||||||||||||||||||||||||
Db    558 GGAACAAGGTTAATGGTCAAACCCCATATCCAGAACCCTGACCCTGCCGTGTACCAGCTG 617

Qy    415 AGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACA 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 AGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACA 677

Qy    475 AATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATG 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 AATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATG 737

Qy    535 AGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 AGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCA 797

Qy    595 TGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAA 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 TGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAA 857

Qy    655 AGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTT 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 AGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTT 917

Qy    715 CAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTG 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 CAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTG 977

Qy    775 CTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCT 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 CTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCT 1037

Qy    835 CCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCT 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 CCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACT----- 1092

Qy    895 ACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTG 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 ------------------------------------------------------------ 1092

Qy    955 GATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 ------------------------------------------------------------ 1092

Qy   1015 CTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTG 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 ------------------------------------------------------------ 1092

Qy   1075 CAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCA 1134
          ||||||||||||||||||||||||||||||||||||
Db   1093 -------------------------CTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCA 1126

Qy   1135 CAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGG 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1127 CAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGG 1186

Qy   1195 GTTTATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGT 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187 GTTTATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGT 1246

Qy   1255 GGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATG 1314
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1247 GGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTATG 1306

Qy   1315 TCCTGCTGCCGATGCCTTCATTAAAAT 1341
          |||||||||||||||||||||||||||
Db   1307 TCCTGCTGCCGATGCCTTCATTAAAAT 1333


RESULT   9
HSTCELA1
LOCUS       HSTCELA1                1131 bp    mRNA    linear   PRI 16-FEB-1995
DEFINITION  H.sapiens mRNA fragment for T-cell receptor alpha chain.
ACCESSION   X01403
VERSION     X01403.1  GI:36752
KEYWORDS    T-cell receptor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1131)
  AUTHORS   Sim,G.K., Yague,J., Nelson,J., Marrack,P., Palmer,E., Augustin,A.
            and Kappler,J.
  TITLE     Primary structure of human T-cell receptor alpha-chain
  JOURNAL   Nature 312 (5996), 771-775 (1984)
  MEDLINE   85086254
COMMENT     Data kindly reviewed (11-FEB-1986) by J. Kappler.
FEATURES             Location/Qualifiers
     source          1. .1131
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             <1. .822
                     /note="putative"
                     /codon_start=1
                     /product="T-cell receptor alpha-chain"
                     /protein_id="CAA25651.1"
                     /db_xref="GI:673422"
                     /translation="IFASLLRAVIASICVVSSMAQKVTQAQTEISVVEKEDVTLDCVY
                     ETRDTTYYLFWYKQPPSGELVFLIRRNSFDEQNEISGRYSWNFQKSTSSFNFTITASQ
                     VVDSAVYFCALDSSASKIIFGSGTRLSIRPNIQNPDPAVYQLRDSKSSDKSVCLFTDF
                     DSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDT
                     FFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS"
     V_region        61. .345
                     /note="putative"
                     /product="variable region of T-cell receptor alpha-chain"
     J_segment       346. .393
```

```
                     /note="putative"
                     /product="joining region of T-cell receptor alpha-chain"
     C_region        394. .744
                     /note="(aa 112-228)"
                     /product="constant region of T-cell receptor alpha-chain"
     misc_feature    745. .804
                     /note="putative transmembrane region"
     misc_feature    805. .819
                     /note="putative cytoplasmic region"
BASE COUNT      286 a    310 c    242 g    293 t
ORIGIN
```

```
  Query Match           50.9%;  Score 683.2;  DB 9;  Length 1131;
  Best Local Similarity 83.1%;  Pred. No. 1.4e-207;
  Matches  853;  Conservative    0;  Mismatches  158;  Indels   15;  Gaps    6;

Qy      98 AAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTACAAATGATTATATCACGTGGT 157
           | || |||| | |   ||||    |  | | |  || ||| | | ||||  |    ||||
Db     107 AGGATGTGACCTTGGACTGTGTGTATGAAACCCGTGATACTACTTATTACTTATTCTGGT 166

Qy     158 ACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTCAAGGATAC---------AAGA 208
           || | ||    || ||   || |       ||  ||||||   |  ||          ||
Db     167 ACAAGCAACCACCAAGTGGAGAATTGGTTTTCCTTATTCGTCGGAACTCTTTTGATGAGC 226

Qy     209 CAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAAAGTCCAGCACTC 268
           |||  |  | ||  |    |  ||   |   |||       ||  |  |||  ||
Db     227 AAAATGAAATAAGTGGTCGGTATTCTTGGAACTTCCAGAAATCCACCAGTTCCTTCAACT 286

Qy     269 TGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCCTCGTGGGTGGTT 328
           | | | |          |   ||| | || || |||| |||     |||   |
Db     287 TCACCATCACAGCCTCACAAGTCGTGGACTCAGCAGTATACTTCTGTGCTCTGGACAGCA 346

Qy     329 CTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGA 388
           |||     |  | | ||||||||| |||||   | | |   |    ||  |||||||||
Db     347 GTGCTTCCAAGATAATCTTTGGATCAGGGACCAGACTCAGCATCCGGCCAAATATCCAGA 406

Qy     389 ACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCC 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407 ACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCC 466

Qy     449 TATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATA 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     467 TATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATA 526

Qy     509 TCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGG 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 TCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGG 586

Qy     569 CCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAG 628
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     587 CCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAG 646

Qy     629 AAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCT 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     647 AAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCT 706

Qy     689 TTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCC 748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     707 TTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCC 766

Qy     749 TCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTG-AGAT 807
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db     767 TCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAAGAT 826

Qy     808 CTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCT--CTGCATTGCCCCTCTT 865
           ||||||| |||||||||||||||||||||||||||||||||||           |||||||||||
Db     827 CTGCAAGACTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTTCTGCATATGCCCCTCTT 886

Qy     866 CTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCT 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     887 CTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCT 946

Qy     926 CTGTGCCCCCCCGGCAAT-GCCACCAACTGGATCCTACCCGAATTTATGATTAAGAT-TG 983
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||| ||
Db     947 CTGTGCCCCCCCGGCAATGGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATGTG 1006

Qy     984 CTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTG 1043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1007 CTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTG 1066

Qy    1044 CCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTC 1103
           |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db    1067 CCTGACATTCACGGCAGAGGCAA-GCTGCTGCAGCCTCCCCTGGCTTTGCACATTCCCTC 1125

Qy    1104 CTGCTC 1109
           ||||||
Db    1126 CTGCTC 1131
```

```
RESULT  10
HSTCARA
LOCUS       HSTCARA                 1064 bp    mRNA    linear   PRI 06-DEC-1992
DEFINITION  H.sapiens mRNA for T-cell antigen receptor alpha-chain.
ACCESSION   X63455
VERSION     X63455.1  GI:36730
KEYWORDS    influenza haemagglutinin peptide specific; T-cell antigen receptor
            (alpha chain); V alpha 1.2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1064)
  AUTHORS   Hewitt,C.R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1991) C.R.A. Hewitt, St. Mary's Hospital Medical
            School, Dept.Of Immunology, Norfolk Place, Paddington, London
            W21PG, UK
REFERENCE   2  (bases 1 to 1064)
  AUTHORS   Hewitt,C.R., Lamb,J.R., Hayball,J., Hill,M., Owen,M.J. and
            O'Hehir,R.E.
  TITLE     Major histocompatibility complex independent clonal T cell anergy
            by direct interaction of Staphylococcus aureus enterotoxin B with
            the T cell antigen receptor
  JOURNAL   J. Exp. Med. 175 (6), 1493-1499 (1992)
  MEDLINE   92268797
FEATURES             Location/Qualifiers
     source          1. .1064
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UB alpha 14/4"
                     /cell_line="HA1.7"
                     /cell_type="T lymphocyte"
     5'UTR           1. .89
     CDS             90. .917
                     /codon_start=1
                     /product="T cell antigen receptor alpha chain"
                     /protein_id="CAA45055.1"
                     /db_xref="GI:36731"
                     /translation="MLLLLVPVLEVIFTLGGTRAQSVTQLGSHVSVSEGALVLLRCNY
                     SSSVPPYLFWYVQYPNQGLQLLLKYTSAATLVKGINGFEAEFKKSETSFHLTKPSAHM
                     SDAAEYFCAVSESPFGNEKLTFGTGTRLTIIPNIQNPDPAVYQLRDSKSSDKSVCLFT
                     DFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPE
                     DTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS"
     sig_peptide     90. .149
     mat_peptide     150. .914
                     /product="T cell antigen receptor alpha chain"
     misc_feature    150. .427
                     /note="V region"
     misc_feature    428. .488
                     /note="J /N region"
     misc_feature    489. .917
                     /note="C region"
     3'UTR           918. .1064
BASE COUNT      258 a    307 c    240 g    259 t
ORIGIN
```

```
Query Match             44.5%;  Score 596.4;  DB 9;  Length 1064;
Best Local Similarity   91.5%;  Pred. No. 1e-179;
Matches  646;  Conservative    0;  Mismatches   51;  Indels    9;  Gaps    1;

Qy    255 AAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTG 314
          ||  |||  |   ||||    |||   |       |||||||| | |||| ||||| |||
Db    359 AACCTCCTTCCACCTGACGAAACCCTCAGCCCATATGAGCGACGCGGCTGAGTACTTCTG 418

Qy    315 CCTCGTGGGTGGTTCTGCA---------AGGCAACTGACCTTTGGATCTGGGACACAATT 365
              ||| |||  ||| ||           | || | ||||||||  |||| |||  | |
Db    419 TGCTGTGAGTGAGTCTCCATTTGGAAATGAGAAATTAACCTTTGGGACTGGAACAAGACT 478

Qy    366 GACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAA 425
          ||  |  ||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    479 CACCATCATACCCAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAA 538

Qy    426 ATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACA 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 ATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACA 598

Qy    486 AAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGA 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    599 AAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGA 658

Qy    546 CTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    659 CTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC 718

Qy    606 CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGA 665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    719 CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGA 778

Qy    666 TGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTC 725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 TGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTC 838

Qy    726 AGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCT 785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 AGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCT 898

Qy    786 GCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCT 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    899 GCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCT 958

Qy    846 TCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGA 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 TCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGA 1018

Qy    906 GGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAA 951
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1019 GGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAA 1064
```

```
RESULT  11
HUMTCAYE
LOCUS       HUMTCAYE                1080 bp    mRNA    linear   PRI 13-JAN-1995
DEFINITION  Human T-cell receptor rearranged alpha-chain V-region (V-D-J) mRNA,
            complete cds.
ACCESSION   M15565
VERSION     M15565.1  GI:338765
KEYWORDS    T-cell receptor antigen; cell surface glycoprotein.
SOURCE      Human cytotoxic T-lymphocyte, cDNA to mRNA, clone L17Ti-alpha.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Leiden,J.M., Fraser,J.D. and Strominger,J.L.
  TITLE     The complete primary structure of the T-cell receptor genes from an
            alloreactive cytotoxic human T-lymphocyte clone
  JOURNAL   Immunogenetics 24 (1), 17-23 (1986)
  MEDLINE   86276770
```

```
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
     gene            1. .1080
                     /gene="TCRA"
     sig_peptide     109. .189
                     /gene="TCRA"
                     /note="T-cell receptor signal peptide"
     CDS             109. .951
                     /gene="TCRA"
                     /note="T-cell receptor precursor"
                     /codon_start=1
                     /protein_id="AAA60627.1"
                     /db_xref="GI:338766"
                     /db_xref="GDB:G00-120-404"
                     /translation="MAMLLGASVLILWLQPDWVNSQQKNDDQQVKQNSPSLSVQEGRI
                     SILNCDYTNSMFDYFLWYKKYPAEGPTFLISISSIKDKNEDGRFTVFLNKSAKHLSLH
                     IVPSQPGDSAVYFCAAKGAGTASKLTFGTGTRLQVTLDIQNPDPAVYQLRDSKSSDKS
                     VCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNN
                     SIIPEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLR
                     LWSS"
     mat_peptide     190. .948
                     /gene="TCRA"
                     /product="T-cell receptor"
     misc_recomb     468. .469
                     /gene="TCRA"
                     /organism="Homo sapiens"
     misc_recomb     524. .525
                     /gene="TCRA"
                     /organism="Homo sapiens"
BASE COUNT      303 a    274 c    239 g    264 t
ORIGIN      Chromosome 14q11.2.
```

```
Query Match             42.5%;  Score 569.6;  DB 9;  Length 1080;
Best Local Similarity   84.2%;  Pred. No. 4.3e-171;
Matches  670;  Conservative    0;  Mismatches  114;  Indels   12;  Gaps    2;

Qy    141 TGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTCAAGG 200
          |||||||  ||   ||||||  || | |  ||     |||| ||   |||  | ||
Db    276 TGATTATTTCCTATGGTACAAAAAATACCCTGCTGAAGGTCCTACATTCCTGATATCTAT 335

Qy    201 ATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAAAGTC 260
          |      |  || | || ||| ||||  |     |    | ||       ||| ||   |
Db    336 AAGTTCCATTAAGGATAAAAATGAAGATGGAAGATTCACTGTCTTCTTAAACAAAAGTGC 395

Qy    261 CAGCACTCTGAGCCTGCCCCGGGTTTCC------CTGAGCGACACTGCTGTGTACTACTG 314
          ||     ||    |||| |   ||  ||      |   | ||| |||| |||||||  |||
Db    396 CAAGCACCTCTCTCTGCACATTGTGCCCTCCCAGCCTGGAGACTCTGCAGTGTACTTCTG 455

Qy    315 ------CCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGAC 368
                 | ||  ||  |||| ||  |||| ||||||||  |||| |||  | |
Db    456 TGCAGCAAAGGGGGCCGGCACTGCCAGTAAACTCACCTTTGGGACTGGAACAAGACTTCA 515

Qy    369 TGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATC 428
            ||    | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 GGTCACGCTCGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATC 575

Qy    429 CAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAG 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 CAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAG 635

Qy    489 TAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTT 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 TAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTT 695

Qy    549 CAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTT 608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    696 CAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTT 755

Qy    609 CAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGT 668
```

```
Db    756 CAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGT 815

Qy    669 CAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGT 728

Db    816 CAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGT 875

Qy    729 GATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCG 788

Db    876 GATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCG 935

Qy    789 GCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCC 848

Db    936 GCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCC 995

Qy    849 TCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGT 908

Db    996 TCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCCACCCCCAAGGAGGT 1055

Qy    909 GAAAGCTGCTACCACC 924

Db   1056 GAAAGCTGCTACCACC 1071
```

```
RESULT  12
HSTCRAC4
LOCUS       HSTCRAC4                  840 bp    DNA     linear   PRI 12-NOV-1996
DEFINITION  Human DNA for T-cell receptor constant region alpha-chain exon 4.
ACCESSION   X05002 M14861
VERSION     X05002.1  GI:36917
KEYWORDS    constant region; T-cell receptor; T-cell receptor alpha.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   Baer,R., Lefranc,M.P., Minowada,J., Forster,A., Stinson,M.A. and
            Rabbitts,T.H.
  TITLE     Organization of the T-cell receptor alpha-chain gene and
            rearrangement in human T-cell leukaemias
  JOURNAL   Mol. Biol. Med. 3 (3), 265-277 (1986)
  MEDLINE   86284199
COMMENT     See also X02883 for complete TCR C(alpha) gene with some
            differences in exon 4 sequence.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B-cell DNA library"
                     /clone="lamdaDalpha3"
     intron          <1. .34
                     /note="intron III"
     exon            35. .>36
                     /number=4
     misc_feature    553. .558
                     /note="put. polyA signal"
BASE COUNT      175 a    246 c    186 g    233 t
ORIGIN
```

```
  Query Match           40.2%;  Score 538.8;  DB 9;  Length 840;
  Best Local Similarity  97.8%;  Pred. No. 3.1e-161;
  Matches  546;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843

Db     13 CTTCTGTTCTTCCTCATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 72

Qy    844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903

Db     73 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 132

Qy    904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963

Db    133 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACC 192

Qy    964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023

Db    193 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 252

Qy   1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083

Db    253 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 312

Qy   1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143

Db    313 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 372

Qy   1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203

Db    373 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 432

Qy   1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263

Db    433 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 492

Qy   1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323

Db    493 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 552

Qy   1324 CGATGCCTTCATTAAAAT 1341

Db    553 CGATGCCTTCATTAAAAT 570
```

```
RESULT  13
HUMTCACA4
LOCUS       HUMTCACA4                 840 bp    DNA     linear   PRI 13-JAN-1995
DEFINITION  Human T-cell receptor germline alpha-chain C-region gene, exon 4,
            clone lambda-D-alpha-3.
ACCESSION   M14861
VERSION     M14861.1  GI:338712
KEYWORDS    C-region; T-cell receptor; germline.
SEGMENT     4 of 4
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   Baer,R., Lefranc,M.P., Minowada,J., Forster,A., Stinson,M.A. and
            Rabbitts,T.H.
  TITLE     Organization of the T-cell receptor alpha-chain gene and
            rearrangement in human T-cell leukaemias
  JOURNAL   Mol. Biol. Med. 3 (3), 265-277 (1986)
  MEDLINE   86284199
COMMENT     Clean copy of sequence [1] kindly provided by M.-P.Lefranc
            (01-AUG-1986).
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
                     /germline
     gene            join(M14858.1:1. .300,M14859.1:1. .120,M14860.1:1. .240,
                     1. .568)
                     /gene="TCRA"
     intron          <1. .34
                     /gene="TCRA"
                     /note="TCA intron C"
     exon            35. .568
                     /partial
                     /gene="TCRA"
                     /note="G00-120-404"
BASE COUNT      175 a    246 c    186 g    233 t
ORIGIN      About 0.5 kb after segment 3.
```

```
  Query Match             40.2%;  Score 538.8;  DB 9;  Length 840;
  Best Local Similarity   97.8%;  Pred. No. 3.1e-161;
  Matches  546;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843
           || | | | |    ||    ||||||||||||||||||||||||||||||||||||||||||
Db      13 CTTCTGTTCTTCCTCATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 72

Qy     844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 132

Qy     904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 192

Qy     964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 252

Qy    1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 312

Qy    1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 372

Qy    1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 432

Qy    1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     433 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 492

Qy    1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     493 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 552

Qy    1324 CGATGCCTTCATTAAAAT 1341
           ||||||||||||||||||
Db     553 CGATGCCTTCATTAAAAT 570


RESULT  14
CNS01RH3
LOCUS       CNS01RH3               208953 bp    DNA     linear   PRI 07-DEC-2001
DEFINITION  Human chromosome 14 DNA sequence BAC C-2555K7 of library CalTech-D
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL160314
VERSION     AL160314.7  GI:17426585
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208953)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208953)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 8, 2001 this sequence version replaced gi:16945087.
            --------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Downstream BAC (overlapping the SP6 end) : R-137H15 (AC=AL135998)
            --------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 10.49x in Q20 bases; sum-of-contigs
            ---------------
            ------------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :
            10 - 19   :       7
            20 - 29   :      23
            30 - 39   :     108
            40 - 49   :     590
            50 - 59   :    1914
            60 - 69   :    3072
            70 - 79   :    9115
            80 - 89   :   37378
            90 - 99   :  156746
            ------------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .208953
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="C-2555K7"
                     /clone_lib="CalTech-D"
     STS             11348. .11777
                     /note="matching EMBL:M14862
                     RHdb:RH11152
                     dbSTS:STS328
                     Identified using the e-PCR software (G. Schuler)"
     STS             20600. .20768
                     /note="matching EMBL:G13536
                     RHdb:RH13835
                     dbSTS:STS1067
                     Identified using the e-PCR software (G. Schuler)"
     STS             33864. .34045
                     /note="matching EMBL:D12193
                     RHdb:RH96131
                     RHdb:RH1200
                     dbSTS:STS15139
                     Identified using the e-PCR software (G. Schuler)"
     STS             64772. .64875
                     /note="matching EMBL:R42714
                     RHdb:RH53928
                     dbSTS:STS21457
                     Identified using the e-PCR software (G. Schuler)"
     STS             81032. .81213
                     /note="matching EMBL:T50419
                     RHdb:RH91439
                     dbSTS:STS64574
                     Identified using the e-PCR software (G. Schuler)"
     STS             81063. .81213
                     /note="matching EMBL:G14776
                     RHdb:RH8029
                     dbSTS:STS2224
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    60111 a  46575 c  45879 g  56388 t
ORIGIN


  Query Match             39.9%;  Score 535.6;  DB 9;  Length 208953;
  Best Local Similarity   97.5%;  Pred. No. 1.1e-159;
  Matches  544;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;
```

```
Qy     784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843
           || | | | |    ||    ||||||||||||||||||||||||||||||||||||||||
Db   20507 CTTCTGTTCTTCCTCATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 20566

Qy     844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20567 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 20626

Qy     904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20627 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 20686

Qy     964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20687 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 20746

Qy    1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20747 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 20806

Qy    1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20807 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 20866

Qy    1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   20867 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTT 20926

Qy    1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20927 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 20986

Qy    1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   20987 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTATGTCCTGCTGC 21046

Qy    1324 CGATGCCTTCATTAAAAT 1341
           ||||||||||||||||||
Db   21047 CGATGCCTTCATTAAAAT 21064
```

RESULT 15
AC023226
LOCUS AC023226 175053 bp DNA linear HTG 03-MAR-2000
DEFINITION Homo sapiens chromosome 14 clone RP11-576P2 map 14, WORKING DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION AC023226
VERSION AC023226.3 GI:7145013
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175053)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 14, clone RP11-576P2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175053)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:7139832.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L6703
Center clone name: 576_P_2
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161058 bases at least Q40
Consensus quality: 167879 bases at least Q30
Consensus quality: 170671 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 173053; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1065: contig of 1065 bp in length
* 1066 1165: gap of 100 bp
* 1166 2413: contig of 1248 bp in length
* 2414 2513: gap of 100 bp
* 2514 4839: contig of 2326 bp in length
* 4840 4939: gap of 100 bp
* 4940 6965: contig of 2026 bp in length
* 6966 7065: gap of 100 bp
* 7066 8841: contig of 1776 bp in length
* 8842 8941: gap of 100 bp
* 8942 11475: contig of 2534 bp in length
* 11476 11575: gap of 100 bp
* 11576 14560: contig of 2985 bp in length
* 14561 14660: gap of 100 bp
* 14661 19338: contig of 4678 bp in length
* 19339 19438: gap of 100 bp
* 19439 22444: contig of 3006 bp in length
* 22445 22544: gap of 100 bp
* 22545 27238: contig of 4694 bp in length
* 27239 27338: gap of 100 bp
* 27339 33032: contig of 5694 bp in length
* 33033 33132: gap of 100 bp
* 33133 38008: contig of 4876 bp in length
* 38009 38108: gap of 100 bp
* 38109 43595: contig of 5487 bp in length
* 43596 43695: gap of 100 bp
* 43696 51575: contig of 7880 bp in length
* 51576 51675: gap of 100 bp
* 51676 59667: contig of 7992 bp in length
* 59668 59767: gap of 100 bp
* 59768 69201: contig of 9434 bp in length
* 69202 69301: gap of 100 bp
* 69302 79384: contig of 10083 bp in length
* 79385 79484: gap of 100 bp
* 79485 91749: contig of 12265 bp in length
* 91750 91849: gap of 100 bp

```
          *    91850   112964: contig of 21115 bp in length
          *   112965 113064: gap of      100 bp
          *   113065   134547: contig of 21483 bp in length
          *   134548 134647: gap of      100 bp
          *   134648   175053: contig of 40406 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175053
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-576P2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1065
                     /note="assembly_fragment"
     misc_feature    1166. .2413
                     /note="assembly_fragment"
     misc_feature    2514. .4839
                     /note="assembly_fragment"
     misc_feature    4940. .6965
                     /note="assembly_fragment"
     misc_feature    7066. .8841
                     /note="assembly_fragment"
     misc_feature    8942. .11475
                     /note="assembly_fragment"
     misc_feature    11576. .14560
                     /note="assembly_fragment"
     misc_feature    14661. .19338
                     /note="assembly_fragment"
     misc_feature    19439. .22444
                     /note="assembly_fragment"
     misc_feature    22545. .27238
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    27339. .33032
                     /note="assembly_fragment"
     misc_feature    33133. .38008
                     /note="assembly_fragment"
     misc_feature    38109. .43595
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    43696. .51575
                     /note="assembly_fragment"
     misc_feature    51676. .59667
                     /note="assembly_fragment"
     misc_feature    59768. .69201
                     /note="assembly_fragment"
     misc_feature    69302. .79384
                     /note="assembly_fragment"
     misc_feature    79485. .91749
                     /note="assembly_fragment"
     misc_feature    91850. .112964
                     /note="assembly_fragment"
     misc_feature    113065. .134547
                     /note="assembly_fragment"
     misc_feature    134648. .175053
                     /note="assembly_fragment"
BASE COUNT    49542 a  38101 c  38584 g  46814 t   2012 others
ORIGIN
```

Query Match 39.8%; Score 534; DB 2; Length 175053;
Best Local Similarity 97.3%; Pred. No. 3.4e-159;
Matches 543; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy    784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843
          || | | | |    ||    |||||||||||||||||||||||||||||||||||||||||
Db  75682 CTTCTGTTCTTCCTCATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 75741

Qy    844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  75742 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 75801

Qy    904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  75802 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACA 75861

Qy    964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  75862 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 75921

Qy   1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  75922 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 75981

Qy   1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  75982 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 76041

Qy   1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||

Db  76042 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCCGGAGAATGTTGTGAGGGGTTTATTTT 76101

Qy   1204 TTTTTAATAGTGTTCATAAAGAAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||

Db  76102 TTTTTAATAGTGTTCATAAAGAAAGACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 76161

Qy   1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  76162 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 76221

Qy   1324 CGATGCCTTCATTAAAAT 1341
          ||||||||||||||||||
Db  76222 CGATGCCTTCATTAAAAT 76239
```

Search completed: June 6, 2002, 07:39:08
Job time: 5371 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:37 ; Search time 2649.09 Seconds
(without alignments)
6832.315 Million cell updates/sec

Title: US-08-881-509-1
Perfect score: 1341
Sequence: 1 ATGAGGCAAGTGGCGAGAGT.....GCCGATGCCTTCATTAAAAT 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	972.6	72.5	1468	11	BC020270	BC020270 Homo sapi
	2	768.6	57.3	1012	10	BM479883	BM479883 AGENCOURT
	3	730.6	54.5	834	10	BI911817	BI911817 603065260
	4	722.8	53.9	869	10	BI764928	BI764928 603047318
	5	704.6	52.5	929	10	BG684892	BG684892 602636627
c	6	694.4	51.8	822	9	AI816828	AI816828 wj34f05.x
	7	688.4	51.3	967	10	BI911009	BI911009 603068729
	8	651.8	48.6	717	10	BI772543	BI772543 603054462
	9	632.8	47.2	1054	10	BM457321	BM457321 AGENCOURT
c	10	620.8	46.3	913	10	BI524586	BI524586 603052009
	11	611.4	45.6	973	10	BG719964	BG719964 602691317
	12	603	45.0	871	10	BG756578	BG756578 602713728
c	13	593.2	44.2	638	9	AW514300	AW514300 hd74f09.x
c	14	584	43.5	749	9	AI351789	AI351789 qt09c03.x
c	15	578	43.1	621	9	AW410870	AW410870 fh08h05.y
c	16	577	43.0	617	9	AW512331	AW512331 xx73c07.x
c	17	570.8	42.6	610	9	AW976912	AW976912 EST389021
c	18	563.6	42.0	618	9	AW026377	AW026377 wv12e05.x
c	19	556.4	41.5	619	9	AI660195	AI660195 we61g07.x
c	20	555.6	41.4	721	9	AI357224	AI357224 qx17h02.x
c	21	551.8	41.1	692	9	AI471656	AI471656 tl99a09.x
	22	547.6	40.8	762	10	BG758732	BG758732 602712974
c	23	534.4	39.9	589	9	AW263612	AW263612 xq46h02.x
c	24	534	39.8	582	9	AI813335	AI813335 wj33e12.x
	25	530.8	39.6	595	10	BI907181	BI907181 603065776
	26	516.4	38.5	887	10	BM457146	BM457146 AGENCOURT
c	27	510.8	38.1	558	9	AI741470	AI741470 wg11b02.x
	28	500	37.3	908	10	BI771427	BI771427 603059463
c	29	471.4	35.2	551	9	AW130991	AW130991 xe68c04.x
	30	471.2	35.1	599	10	BE254521	BE254521 601115131
c	31	470.8	35.1	527	9	AA854890	AA854890 aj78e02.s
c	32	464.8	34.7	485	9	AI094938	AI094938 qa18d09.x
	33	464.6	34.6	756	10	BG393458	BG393458 602411822
c	34	462.8	34.5	518	9	AW591225	AW591225 xo12g01.x
c	35	462.6	34.5	565	9	AA400994	AA400994 zu50f04.s
c	36	462	34.5	505	9	AA916904	AA916904 oh85a07.s
c	37	460.8	34.4	511	9	AI831179	AI831179 wj81h09.x
c	38	460.6	34.3	491	9	AW450946	AW450946 UI-H-BI3-
	39	456.8	34.1	554	10	BG257692	BG257692 602377574
c	40	446.2	33.3	495	9	AI091979	AI091979 oz96f11.x
c	41	445.2	33.2	488	9	AA284727	AA284727 zt24b05.s
	42	440.4	32.8	800	10	BG826688	BG826688 602750823
	43	440.4	32.8	896	10	BI833474	BI833474 603088138
	44	438.4	32.7	569	10	BE258557	BE258557 601108335
c	45	437	32.6	458	9	AI718656	AI718656 as94g02.x

ALIGNMENTS

RESULT 1
BC020270
LOCUS BC020270 1468 bp mRNA linear HTC 19-DEC-2001
DEFINITION Homo sapiens, clone IMAGE:4764815, mRNA.
ACCESSION BC020270
VERSION BC020270.1 GI:17939628
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1468)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 40 Row: h Column: 7
This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein
This clone has the following problem: no cloning site / microdeletion.

```
FEATURES             Location/Qualifiers
     source          1. .1468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4764815"
                     /tissue_type="Primary B-Cells from Tonsils"
                     /clone_lib="NIH_MGC_48"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      392 a    362 c    329 g    385 t
ORIGIN
```

Query Match 72.5%; Score 972.6; DB 11; Length 1468;
Best Local Similarity 97.6%; Pred. No. 1.9e-254;
Matches 987; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy    331 GCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAAC 390
Db    421 GGAAACAAACTCACCTTTGGGACAGGCACTCAGCTAAAAGTGGAACTCAATATCCAGAAC 480

Qy    391 CCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTA 450
Db    481 CCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTA 540

Qy    451 TTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATC 510
Db    541 TTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATC 600

Qy    511 ACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCC 570
Db    601 ACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCC 660

Qy    571 TGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAA 630
Db    661 TGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAA 720

Qy    631 GACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTT 690
Db    721 GACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTT 780

Qy    691 GAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTC 750
Db    781 GAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTC 840

Qy    751 CTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTG 810
Db    841 CTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTG 900

Qy    811 CAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCC 870
Db    901 CAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCC 960

Qy    871 TCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTG 930
Db    961 TCTCCAAACAGAGGGAACTCTCCCACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTG 1020

Qy    931 CCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGA 990
Db   1021 CCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGA 1080

Qy    991 GCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACA 1050
Db   1081 GCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACA 1140

Qy   1051 TTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCC 1110
Db   1141 TTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCC 1200

Qy   1111 CCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCT 1170
Db   1201 CCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCT 1260

Qy   1171 AGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAATAC 1230
Db   1261 AGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAATAC 1320

Qy   1231 ATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATG 1290
Db   1321 ATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATG 1380

Qy   1291 CTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1341
Db   1381 CTGTGTATCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1431
```

```
RESULT   2
BM479883
LOCUS       BM479883                1012 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6464798 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5577362
            5', mRNA sequence.
ACCESSION   BM479883
VERSION     BM479883.1  GI:18528925
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12330  row: 1  column: 03
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5577362"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      251 a    281 c    220 g    256 t      4 others
ORIGIN
```

Query Match 57.3%; Score 768.6; DB 10; Length 1012;
Best Local Similarity 94.6%; Pred. No. 8.8e-199;
Matches 826; Conservative 0; Mismatches 43; Indels 4; Gaps 3;

```
Qy    379 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 438
Db    143 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 202

Qy    439 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 498
Db    203 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 262
```

```
Qy     499 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     263 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 322

Qy     559 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     323 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 382

Qy     619 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     383 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 442

Qy     679 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     443 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 502

Qy     739 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     503 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 562

Qy     799 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     563 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 622

Qy     859 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCT 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     623 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCT 682

Qy     919 ACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAA 978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     683 ACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAA 742

Qy     979 GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGT 1038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     743 GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGT 802

Qy    1039 CACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATT 1098
           |||||||||||| ||||||||||||| |          ||||||||||||||||||||||
Db     803 CACTGCCTGACA-TCACGGCAGAGGCAGGCTGCTGCAGCCCTCCCCTGGCTGTGCACATT 861

Qy    1099 CCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTT 1158
           ||||  ||| ||||||||||||  |||||||| ||||||||||||| ||||||||||
Db     862 TCCTCTGGCT-CCCAGAGACTGCCNTCGCCATCCAACAGATGATGGATCCTCAGTGGGTT 920

Qy    1159 CTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTC 1218
           ||||||||||||||||    | |   |  ||| ||||||||||||||| |||| ||
Db     921 CTCTTGGGCTCTAGGTCTGCAAAATGTTNNTGAAGGGTTTATTTTTTTTTTTATAGGGT-- 978

Qy    1219 ATAAAGAAATACATAGTATTCTTCTTCTCAAGA 1251
              ||  |||   |||||||| |  ||||| ||
Db     979 CNTAAAGAATTACTAGTATTCCTTCTCTCACGA 1011
```

```
RESULT   3
BI911817
LOCUS       BI911817                 834 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  603065260F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5214375 5',
            mRNA sequence.
ACCESSION   BI911817
VERSION     BI911817.1  GI:16175689
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11538  row: g  column: 16
            High quality sequence start: 26
            High quality sequence stop: 828.
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5214375"
                     /clone_lib="NIH_MGC_118"
                     /tissue_type="leukocyte"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
                     (destroyed); RNA source leukocytes from anonymous pool of
                     non-activated  adult donors.  Library is oligo-dT primed
                     and directionally cloned  (EcoRV site is destroyed upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     1.2-3.3 kb.  Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking code 027. Note:
                     this is a NIH_MGC Library."
BASE COUNT       214 a    244 c    174 g    202 t
ORIGIN
```

```
  Query Match           54.5%;  Score 730.6;  DB 10;  Length 834;
  Best Local Similarity 96.2%;  Pred. No. 2e-188;
  Matches  759;  Conservative    0;  Mismatches   29;  Indels    1;  Gaps    1;

Qy     333 AAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCC 392
           | |  | || | |||||||   || ||  || |  |||||  |||  ||||||||||||
Db      36 AGGAAATCTCATCTTTGGAAAAGGCACTAAACTCTCTGTTAAACCAAATATCCAGAACCC 95

Qy     393 TGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATT 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 TGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATT 155

Qy     453 CACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCAC 512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 CACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCAC 215

Qy     513 AGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTG 572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 AGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTG 275

Qy     573 GAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGA 632
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     276 GAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGA 335

Qy     633 CACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGA 692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     336 CACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGA 395

Qy     693 AACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCT 752
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     396 AACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCT 455

Qy     753 GAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCA 812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     456 GAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCA 515

Qy     813 AGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCCTCTTCTCCCTC 872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     516 AGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCCTCTTCTCCCTC 575

Qy     873 TCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCC 932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     576 TCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCC 635
```

```
Qy     933 CCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGC 992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     636 CCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGC 695

Qy     993 TGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATT 1052
           |||||||||||||||  ||||    | ||||||||||||||||||||||||||||||||||
Db     696 TGCCAAACACTGCTGCACCCCCTCTGTTCCCCTTATTGCTGCTTGTCACTGCCTGACATT 755

Qy    1053 CACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCC 1112
           |||||||||||||||||||||||||  |||||||||||||||||| ||||||||||||||
Db     756 CACGGCAGAGGCAAGGCTGCTGCAGCTCCCCCTGGCTGTGCACATT-CCTCCTGCTCCCC 814

Qy    1113 AGAGACTGC 1121
           |||||||||
Db     815 AGAGACTGC 823
```

```
RESULT   4
BI764928
LOCUS       BI764928                869 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603047318F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187413 5',
            mRNA sequence.
ACCESSION   BI764928
VERSION     BI764928.1  GI:15756506
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11468  row: d  column: 06
            High quality sequence stop: 763.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5187413"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      228 a    239 c    188 g    214 t
ORIGIN
```

```
  Query Match           53.9%;  Score 722.8;  DB 10;  Length 869;
  Best Local Similarity  95.2%;  Pred. No. 2.7e-186;
  Matches  811;  Conservative    0;  Mismatches   32;  Indels    9;  Gaps    6;

Qy     422 CTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGT 481
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGT 60
```

```
Qy     482 CACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTA 541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTA 120

Qy     542 TGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAA 601
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     121 TGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTGGCATGTGCAA 180

Qy     602 ACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCT 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCT 240

Qy     662 GTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACC 721
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACC 300

Qy     722 TGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGA 781
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGA 360

Qy     782 CGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGC 841
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGC 420

Qy     842 TCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCA 901
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCA 480

Qy     902 AGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTA 961
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTA 540

Qy     962 CCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTT 1021
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     541 CCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGACACCCCCTCTGTT 600

Qy    1022 CCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCT- 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     601 CCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTC 660

Qy    1081 CCCCTGGCTGTGCACATTCCCTCCT-GCTCCCCAGAGACTGCCTCCGCCATCCCACAGAT 1139
           |||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||
Db     661 CCCCTGGCTGTGCACATTCCCTCCTGGCTCCCCAGAGACTGCCTCCGACATCCCACAGAT 720

Qy    1140 GATG-GATCTTCAGTGGGTT--CTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGT 1196
           |||| |||||||||| || |    |||||||||||||||| || |   | || |||
Db     721 GATGCGATCTTCAGTAGGGTACTACTTGGGCTCTAGGTCCTGCAGGAATGTTGGAAGGGG 780

Qy    1197 TTATTTTTTTTTAATAGTGTTCAT-AAAGAAATACATAG---TATTCTTCTTCTCAAGAC 1252
           ||| |    | |  |||||||| ||||||||||||||   || |    | ||||||||
Db     781 TTAATAAAAATAATAAGTGTTCATGAAAGAAATACATAGGAATACTAAATTACTCAAGAC 840

Qy    1253 GTGGGGGGAAAT 1264
           |||||||| |||
Db     841 GTGGGGGGCAAT 852
```

```
RESULT   5
BG684892
LOCUS       BG684892                929 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602636627F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4764295 5',
            mRNA sequence.
ACCESSION   BG684892
VERSION     BG684892.1  GI:13916289
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1621  row: f  column: 08
             High quality sequence stop: 840.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4764295"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       225 a    247 c    220 g    237 t
ORIGIN
```

Query Match 52.5%; Score 704.6; DB 10; Length 929;
Best Local Similarity 91.9%; Pred. No. 2.6e-181;
Matches 788; Conservative 0; Mismatches 64; Indels 5; Gaps 4;

```
Qy    338 AACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACC 397
          || ||| ||| ||   ||||||   | ||   ||    ||||||||||||||||||||||
Db     77 AATTGATCTTCGGGAGTGGGACCAGACTGCTGGTCAGGCCTGATATCCAGAACCCTGACC 136

Qy    398 CTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCG 457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 CTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCG 196

Qy    458 ATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACA 517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 ATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACA 256

Qy    518 AAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCA 577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 AAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCA 316

Qy    578 ACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCT 637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 ACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCT 376

Qy    638 TCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAG 697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 TCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAG 436

Qy    698 ATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAG 757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAG 496

Qy    758 TGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATT 817
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 TGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATT 556

Qy    818 GTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAA 877
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 GTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAA 616
```

```
Qy    878 ACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCC 937
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    617 ACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG-TACCACCTCTGTGCCCCCCC 675

Qy    938 GGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCA 997
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    676 GGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTG-CA 734

Qy    998 AACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGG 1057
          ||||||||||| ||||     | || |  | |  |||||||||| ||||||||||||
Db    735 AACACTGCTGCACCCCCTCTGTTCCCTTATTGGTCGCTTGTCACTGGCTGACATTCACGG 794

Qy   1058 CAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGA 1117
          |||||||||||| || ||||||||||||||| |||||||||| ||  |     ||||||||
Db    795 CAGAGGCAAGGTTGTTGCAGCCTCCCCTGGGTGTGCACATTTCCCTCCTGCTCCCAGAGA 854

Qy   1118 CTGCCTCCGCCATCCCACAGATGATGGATCTTCA-GTGGGTTCTCTTGGGCTCTAGGTCC 1176
            ||   |  ||  |||||||||||||||||||| | ||||||||||||||    | |
Db    855 TGGC--TCCGCAATCCACAGATGATGGATCTTCAGGAGGGTTCTCTTGGGCTTCTAGGTC 912

Qy   1177 TGGAGAATGTTGTGAGG 1193
            |  ||||| ||  ||
Db    913 CTGGAAATGTCGTACGG 929
```

```
RESULT   6
AI816828/c
LOCUS        AI816828                 822 bp    mRNA    linear   EST 21-DEC-1999
DEFINITION   wj34f05.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2404737 3'
             similar to gb:M12959 T-CELL RECEPTOR ALPHA CHAIN C REGION (HUMAN);,
             mRNA sequence.
ACCESSION    AI816828
VERSION      AI816828.1  GI:5435907
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 822)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 701    Std Error: 0.00
             Seq primer: -40UP from Gibco
             High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2404737"
                     /clone_lib="NCI_CGAP_Kid12"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
```

Fatima Bonaldo. "

BASE COUNT 218 a 179 c 228 g 194 t 3 others
ORIGIN

Query Match 51.8%; Score 694.4; DB 9; Length 822;
Best Local Similarity 95.9%; Pred. No. 1.5e-178;
Matches 765; Conservative 0; Mismatches 28; Indels 5; Gaps 5;

```
Qy    547 TTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCC 606
          ||||||||  ||  | ||||||  |||||||||||||||||||| ||||||||  ||||||||
Db    819 TTCAAGAGGCACCATTCTGTGGGCTGGAGCAACAAATCTGAC-TTGCATGTTCAAACGCC 761

Qy    607 TTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC-TGTGA 665
          |||||||||||||||||||||||||||| |  | |||   ||||||||||||| |||||
Db    760 -TCAACAACAGCATTATTCCAGAAGACACTTCTTCCCCAGCCCCAGAAAGTTCCNTGTGA 702

Qy    666 TGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTC 725
          |||||||||||||||||||||||||||| |||||||| | ||| ||||||||||||||||
Db    701 TGTCAAGCTGGTCGAGAAAAGCTTTGAACCAGATACGACCNTAACCTTTCAAAACCTGTC 642

Qy    726 AGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCT 785
          |||||||||| |||||||| |||||||||| | ||||||||||||||||||||||||
Db    641 AGTGATTGGGTCCCGAATCCTCTTCCTGAAAGTGCCGGGGTTTAATCTGCTCATGACGCT 582

Qy    786 GCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCT 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 GCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCT 522

Qy    846 TCCTCTGCATTGCCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGA 905
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 TCCTNTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGA 462

Qy    906 GGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACCCG 965
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    461 GGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACACG 402

Qy    966 AATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCT 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 AATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCT 342

Qy    1026 TATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCT 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 TATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCT 282

Qy    1086 GGCTGTGCACATTCCCTCCTGCT-CCCCAGAGACTGCCTCCGCCATCCCACAGATGATGG 1144
           |||||||||||||||||||||||| ||||| ||| |||||||||||||||||||||||||
Db     281 GGCTGTGCACATTCCCTCCTGCTCCCCCAAAGAATGCCTCCGCCATCCCACAGATGATGG 222

Qy    1145 ATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTA-TTTT 1203
           |||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Db     221 ATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCCGGAGAATGTTGTGAGGGGTTTATTTTT 162

Qy    1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     161 TTTTTAATAGTGTTCATAAAGAAAGACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 102

Qy    1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 42

Qy    1324 CGATGCCTTCATTAAAAT 1341
           ||||||||||||||||||
Db      41 CGATGCCTTCATTAAAAT 24
```

RESULT 7
BI911009
LOCUS BI911009 967 bp mRNA linear EST 16-OCT-2001
DEFINITION 603068729F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5217870 5', mRNA sequence.
ACCESSION BI911009
VERSION BI911009.1 GI:16174524
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 967)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11547 row: i column: 07
High quality sequence start: 3
High quality sequence stop: 746.

FEATURES Location/Qualifiers
source 1. .967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5217870"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

BASE COUNT 240 a 306 c 216 g 205 t
ORIGIN

Query Match 51.3%; Score 688.4; DB 10; Length 967;
Best Local Similarity 96.5%; Pred. No. 7e-177;
Matches 747; Conservative 0; Mismatches 21; Indels 6; Gaps 4;

```
Qy     534 GAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTT-TG 592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       2 GAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTGGG 61

Qy     593 CATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAG 652
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 CATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAG 121

Qy     653 AAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACT 712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 AAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACT 181

Qy     713 TTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATC 772
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 TTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATC 241

Qy     773 TGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTG 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 TGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTG 301

Qy     833 CTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTC 892
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTC 361

Qy     893 CTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAAC 952
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAAC 421
```

```
Qy    953 TGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACC 1012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    422 TGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCAAC 481

Qy   1013 CCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGG-CAGAGGCAAGGCTG 1071
          |||||||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db    482 CCCTCTGTTCCCTTATTGCTGCTTGTCACTGGCTGACATTCACGGCCAGAGGCAAGGCTG 541

Qy   1072 CTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATC 1131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATC 601

Qy   1132 CCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGA 1191
          |||||||||||||||||||||||||| |||||||||||||||||| |||||| |||||
Db    602 CCACAGATGATGGATCTTCAGTGGGTCCTCTTGGGCTCTAGGTCCTGCAGAATG-TGTGA 660

Qy   1192 GGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGA 1251
          |||||| || |   |   | |   |||| ||||||||||||||||||||||||||||||
Db    661 GGGGTTCATCTCCCTAATAGTG---TCATTAAGAAATACATAGTATTCTTCTTCTCAAGA 717

Qy   1252 CGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCG 1305
          |||||||||||| |||||||||||||||| |||||| ||||||| |||| |||
Db    718 CGTGGGGGGAAACTATCTCATTATCGAGGCACTGCTAAGCTGTGTATCTGAGCG 771
```

RESULT 8
BI772543
LOCUS BI772543 717 bp mRNA linear EST 25-SEP-2001
DEFINITION 603054462F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204103 5', mRNA sequence.
ACCESSION BI772543
VERSION BI772543.1 GI:15764121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 717)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11511 row: k column: 16
High quality sequence stop: 717.
FEATURES Location/Qualifiers
source 1. .717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5204103"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."
BASE COUNT 170 a 193 c 162 g 192 t
ORIGIN

```
  Query Match             48.6%;  Score 651.8;  DB 10;  Length 717;
  Best Local Similarity   99.4%;  Pred. No. 6.3e-167;
  Matches  675;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy    664 GATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTG 723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTG 60

Qy    724 TCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACG 783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACG 120

Qy    784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 180

Qy    844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 240

Qy    904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACC 300

Qy    964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 360

Qy   1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 420

Qy   1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 480

Qy   1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    481 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTT 540

Qy   1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 600

Qy   1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCT-GGGCGTGTTGTATGTCCTGCTG 1322
          |||||||||||||||| |||||||||||||| ||| ||||||||||||||||||||||||
Db    601 TTATCTCATTATCGAGG-CCTGCTATGCTGTGTATCTGGGGCGTGTTGTATGTCCTGCTG 659

Qy   1323 CCGATGCCTTCATTAAAAT 1341
          ||||||||||||||||||
Db    660 CCGATGCCTTCATTAAAAT 678
```

RESULT 9
BM457321
LOCUS BM457321 1054 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6406850 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583616 5', mRNA sequence.
ACCESSION BM457321
VERSION BM457321.1 GI:18506361
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1054)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM12346  row: p  column: 17
           High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1. .1054
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583616"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      257 a    308 c    237 g    251 t      1 others
ORIGIN

  Query Match             47.2%;  Score 632.8;  DB 10;  Length 1054;
  Best Local Similarity   97.8%;  Pred. No. 1.1e-161;
  Matches  673;  Conservative    0;  Mismatches   12;  Indels    3;  Gaps    3;

Qy     379 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 294

Qy     439 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 354

Qy     499 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 414

Qy     559 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 474

Qy     619 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 534

Qy     679 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 594

Qy     739 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 654

Qy     799 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 714

Qy     859 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTA-CCCCCAAGGAGGTGAAAGCTGC 917
           |||||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||
Db     715 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCCACCCCCCCCAGGAGGTGAAAGCTGC 774

Qy     918 TACCACCTCTGTG-CCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 976
           |||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db     775 TACCACCTCTGTGCCCCCCCCCGGGAATGCCACCAACTGGATCCTACCCGAATTTATGATT 834

Qy     977 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTG-TTCCCTTATTGCTGCT 1035
           |||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||
Db     835 AAGATTGCTGAAGAGCTGCCAAACACTGCTTGCACCCCCTCTGTTTCCCTTATTGCTGCT 894

Qy    1036 TGTCACTGCCTGACATTCACGGCAGAGG 1063
           |||||||||||    | |||||||| | |
Db     895 TGTCACTGCCCTGAAATCACGGCAAAAG 922


RESULT  10
BI524586/c
LOCUS       BI524586                913 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  6030520097T1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201547 3',
            mRNA sequence.
ACCESSION   BI524586
VERSION     BI524586.1  GI:15349378
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11505  row: a  column: 04
           High quality sequence start: 22
           High quality sequence stop: 773.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201547"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      244 a    197 c    264 g    208 t
ORIGIN

  Query Match             46.3%;  Score 620.8;  DB 10;  Length 913;
  Best Local Similarity   92.2%;  Pred. No. 2e-158;
  Matches  841;  Conservative    0;  Mismatches   52;  Indels   19;  Gaps   17;

Qy     405 GTACCAGCTGAGAGACTCTAAATCCAGTGA-CAAGTCTGTCTGCCTATTC-ACCGATTTT 462
           ||||   |||||||||   ||||||| || || |||   |||||||||| |||||||
Db     910 GTACAGCTTGAGAGACTTCTAATCCAGGGACCACGTCGTCCTGCCTATTCGCCCGATTTC 851

Qy     463 GATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT-GATGTGTATATCACAGACAAAAC 521
           |||||||| |||||||||||| ||||||| ||||| || | ||||||||||||||  ||
Db     850 GATTCTCACACAAATGTGTCACCAAGTAAGAATTCTCGACGGGTATATCACAGACATCAC 791

Qy     522 -TGTGCTAG-ACATGAGGTCTATGGACTTCAA---GAGCAACAGTGCTGTGGCCTGGAGC 576
           ||||||||  |||||||||||||||||||||||   |          |||||||||| |||||
Db     790 TTGTGCTAGAACATGAGGTCTATGGACTTCAAAGCGCAACCAGTTGCTGTGGCCAGGAGC 731

Qy     577 AACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATT-CCAGAAGACA- 634
           ||| ||| ||||||||||||||  ||   | |  |||||||||||||| ||||||||||
Db     730 -ACATATCGGACTTTGCATGTGGCAAACGCCTTCACAACAGCATTATTCCCAGAAGACAC 672
```

```
Qy    635 CCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGT-CAAGCTGGTCGAGAAAAGCTTTGAA 693
          ||||||||| ||||||||||||||||||||||||| ||||||||||||||||||| ||||
Db    671 CCTTCTTCCTCAGCCCAGAAAGTTCCTGTGATGTACAAGCTGGTCGAGAAAAGCTCTGAA 612

Qy    694 ACAGATACGAACCTAAAC-TTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCT 752
          |||||| |||||||||| |||| ||||||||||||||||||||||||||||||||||||
Db    611 ACAGATTCGAACCTAAACGTTTCCAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCT 552

Qy    753 GAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCA 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 GAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCA 492

Qy    813 AGATTGTAAGACAGCC-TGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCT 871
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    491 AGATTGTAAGACAGCCTTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCT 432

Qy    872 CTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTG-AAAGCTGCTACCACCTCTGTG 930
          ||||||||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db    431 CTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGCAATGCTGCTACCACCTCTGTG 372

Qy    931 CCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTG-AAG 989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    371 CCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGTAAG 312

Qy    990 AGCTGCCAAAC-ACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGA 1048
          |||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db    311 AGCTGCCAAACTACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTGGTCACTGCCTGA 252

Qy   1049 CATTCACGGCAGAGGCAAGGCTGCTGCAGCCTC-CCCTGGCTGTGCACATTCCCTCCTGC 1107
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    251 CATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCTCCCTGGCTGTGCACATTCCCTCCTGC 192

Qy   1108 TCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGC 1167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 TCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGC 132

Qy   1168 TCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAA 1227
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 TCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAA 72

Qy   1228 TACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCT 1287
          ||||||||||||||||||||||||||||||||||||||||||||||||||   ||| ||
Db     71 TACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGA-TACCTCCT 13

Qy   1288 ATGCTGTGTGTC 1299
          | ||||||| ||
Db     12 AGGCTGTGTATC 1
```

RESULT 11
BG719964

LOCUS	BG719964 973 bp mRNA linear EST 08-MAY-2001
DEFINITION	602691317F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4823332 5', mRNA sequence.
ACCESSION	BG719964
VERSION	BG719964.1 GI:13999151
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 973)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10733 row: b column: 05 High quality sequence stop: 725.

```
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4823332"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation). Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT       243 a    276 c    215 g    239 t
ORIGIN
```

Query Match 45.6%; Score 611.4; DB 10; Length 973;
Best Local Similarity 93.3%; Pred. No. 7.5e-156;
Matches 695; Conservative 0; Mismatches 41; Indels 9; Gaps 5;

```
Qy    372 TTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAG 431
          | | || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 TCTGCCAGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAG 192

Qy    432 TGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAA 491
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 TGACAACTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAA 252

Qy    492 GGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAA 551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 GGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAA 312

Qy    552 GAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAA 611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAA 372

Qy    612 CAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAA 671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 CAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAA 432

Qy    672 GCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGAT 731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 GCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGAT 492

Qy    732 TGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCT 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 TGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCT 552

Qy    792 GTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCT 851
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    553 GTGGTCCAGCTGAGATCTGCAAGATCGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCT 612

Qy    852 GCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAA 911
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    613 GCATTGCCCCTCTTCTCGCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAA 672

Qy    912 AGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAA---T 968
          |||||||||||||||||||||||| | |||||||||||||||||||||||||||||   |
Db    673 AGCTGCTACCACCTCTGTGCCCCCC--GGAATGCCACCAACTGGATCCTACCCGAAATCT 730

Qy    969 TTATGATTAAGATTGCTGAAGAGCTGCCAAA-CACTGCTGCCACCCCCTCTGTTCCCTTA 1027
          |   |        |||||||||||| |||| |||||||| || || ||||| |||||||
Db    731 CTGACTTACGATCGCCTGAAGAGCTGACAAACCACTGCTGGCAACCACTCTGATCCCTTA 790
```

```
Qy   1028 TTGCTGCTTGTCACTGCCTGACATTCACGG-CAGAGGCAAGGCTGCTGCAGCCTCCCCTG 1086
          | |||||  ||||     |||||||||||||| |||||||| ||||||| |||   |||||
Db    791 TCGCTGCATGTCCACTGCTGACATTCACGGCCAGAGGCAACGCTGCTGAAGC--TCCCTG 848

Qy   1087 GCTGTGCACATTCCCTCCTGCTCCC 1111
          | |||   ||||||||| || ||||
Db    849 GTTGTTGCCATTCCCTCATGTTCCC 873
```

```
RESULT  12
BG756578
LOCUS       BG756578                  871 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602713728F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853814 5',
            mRNA sequence.
ACCESSION   BG756578
VERSION     BG756578.1  GI:14067231
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1700  row: h  column: 07
            High quality sequence stop: 828.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4853814"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      216 a    239 c    196 g    220 t
ORIGIN
```

```
  Query Match             45.0%;  Score 603;  DB 10;  Length 871;
  Best Local Similarity   93.1%;  Pred. No. 1.4e-153;
  Matches  710;  Conservative    0;  Mismatches   40;  Indels   13;  Gaps    7;

Qy     322 GGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACC---- 377
           |||||| || |||   | || | ||||||| |||| || |   |  ||||    ||
Db     112 GGTGGTGCTACAAACAAGCTCATCTTTGGAACTGGCACTCTGCTTGCTGTCCAGCCAAGT 171

Qy     378 TGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAA 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172 ACATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAA 231

Qy     438 GTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTC 497
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     232 GTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTC 291
```

```
Qy     498 TGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAA 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     292 TGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAA 351

Qy     558 CAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAG 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     352 CAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAG 411

Qy     618 CATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGT 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     412 CATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGT 471

Qy     678 CGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTT 737
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     472 CGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTT 531

Qy     738 CCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTC 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     532 CCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTC 591

Qy     798 CAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTG 857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     592 CAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGGATTG 651

Qy     858 CCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGC 917
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     652 -CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGC 710

Qy     918 TACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATTA 977
           |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db     711 TACCACCTCTGTG-CCCCCCGGCAATGCCACCAACTGGATCCTACCCGAA-TTATGATTA 768

Qy     978 AGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTG 1037
           |||||||||||||||||||||||     |    |||||||||||||||||||||| |||
Db     769 AGATTGCTGAAGAGCTGCCAAAC----TGGTGCACCCCTCTGTTCCCTTATTGCTG-TTG 823

Qy    1038 TCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCT 1080
           |||||||||||  ||||||||||||||  |  |  ||  |||
Db     824 TCACTGCCTGA-ATTCACGGCAGAGCAAGCTGTTGAGCTCCCT 865
```

```
RESULT  13
AW514300/c
LOCUS       AW514300                  638 bp    mRNA    linear   EST 03-MAR-2000
DEFINITION  hd74f09.x1 NCI_CGAP_Lu28 Homo sapiens cDNA clone IMAGE:2915273 3'
            similar to gb:M12959 T-CELL RECEPTOR ALPHA CHAIN C REGION (HUMAN);,
            mRNA sequence.
ACCESSION   AW514300
VERSION     AW514300.1  GI:7152306
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1. .638
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2915273"
                     /clone_lib="NCI_CGAP_Lu28"
                     /tissue_type="two pooled squamous cell carcinomas"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies."
BASE COUNT       171 a    130 c    200 g    137 t
ORIGIN


  Query Match             44.2%;  Score 593.2;  DB 9;  Length 638;
  Best Local Similarity   95.6%;  Pred. No. 6.2e-151;
  Matches  610;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     610 AACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTC 669
           |||||||    ||| |  ||| |||||||||  ||| ||||  ||| | ||| | ||||||
Db     638 AACAACACATCTATACATGAACACACCTTCACCCCTAGCCACGAATGATCCCGCGATGTC 579

Qy     670 AAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTG 729
           ||||||||||||||   |||||||||||| ||  ||||||| |||||| || ||||||||||
Db     578 AAGCTGGTCGAGACTCGCTTTGAAACACATCAGAACCTACACTTTCTAATCCTGTCAGTG 519

Qy     730 ATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGG 789
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     518 ATTGGGTTCCGAATCCTCCTCCTGACAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGG 459

Qy     790 CTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCT 849
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     458 CTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCT 399

Qy     850 CTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTG 909
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     398 CTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTG 339

Qy     910 AAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATT 969
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     338 AAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATT 279

Qy     970 TATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATT 1029
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     278 TATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATT 219

Qy    1030 GCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCT 1089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     218 GCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCT 159

Qy    1090 GTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTT 1149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158 GTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTT 99

Qy    1150 CAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTA 1209
           ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      98 CAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTA 39

Qy    1210 ATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTC 1247
           |||||||||||||||||||||||||||||||||||||
Db      38 ATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTC 1


RESULT  14
AI351789/c
LOCUS       AI351789                749 bp    mRNA    linear   EST 13-FEB-1999
DEFINITION  qt09c03.x1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1947076 3'
            similar to gb:M12959 T-CELL RECEPTOR ALPHA CHAIN C REGION (HUMAN);,
            mRNA sequence.
ACCESSION   AI351789
VERSION     AI351789.1  GI:4088995
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1496  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 466.
FEATURES                Location/Qualifiers
     source             1. .749
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1947076"
                        /clone_lib="NCI_CGAP_GC4"
                        /tissue_type="pooled germ cell tumors"
                        /lab_host="DH10B"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; 1st strand cDNA was prepared from 3 pooled
                        germ cell tumors, and was then primed with a Not I -
                        oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                        RI adaptors (Pharmacia), digested with Not I and cloned
                        into the Not I and Eco RI sites of the modified pT7T3
                        vector.  Library is normalized. Library was constructed by
                        Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       198 a    176 c    205 g    167 t      3 others
ORIGIN


  Query Match             43.5%;  Score 584;  DB 9;  Length 749;
  Best Local Similarity   92.0%;  Pred. No. 2.1e-148;
  Matches  671;  Conservative    0;  Mismatches   48;  Indels   10;  Gaps    5;

Qy     614 ACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGC 673
           |||||||||||||||||      |||||  ||    ||||||||||||||||||| |||||
Db     749 ACAGCATTATTCCAGANACACCCTTC--GCCAGAGCCAGAAAGTTCCTGTGATG-CAAGC 693

Qy     674 TGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTG 733
           |||| ||||||   ||  | | |           |  ||  | | |||||||||| ||
Db     692 TGGTGGAGAAAGCTTTGAACAGA-----TCGNACCTANCTTTCAGAACTGTCAGTGA-TG 639

Qy     734 GGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGT 793
           |||| || ||||||||| ||| |||||||||||||||||||||||||||||  |
Db     638 GGTTGCGGATCCTCCTCGTGAGAGTGGCCGGGTTTAATCTGCTCATGACGCTCGGCTGTG 579

Qy     794 GGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGC 853
           |  |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||
Db     578 GTCCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCGCCTTCCTGTGC 519

Qy     854 ATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAG 913
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     518 ATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCCACCCCCAAGGAGGTGAAAG 459

Qy     914 CTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATG 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     458 CTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATG 399

Qy     974 ATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTG 1033
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     398 ATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTG 339

Qy    1034 CTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGC-TGCTGCAGCCTCCCCTGGCTGTG 1092
```

```
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    338 CTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCGTGCTGCAGCCTCCCCTGGCTGTG 279

Qy   1093 CACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAG 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 CACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAG 219

Qy   1153 TGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATA 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 TGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATA 159

Qy   1213 GTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCAT 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 GTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCAT 99

Qy   1273 TATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTT 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 TATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTT 39

Qy   1333 CATTAAAAT 1341
          |||||||||
Db     38 CATTAAAAT 30
```

RESULT 15
AW410870/c
LOCUS AW410870 621 bp mRNA linear EST 29-JUN-2000
DEFINITION fh08h05.y1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2961657 3', mRNA sequence.
ACCESSION AW410870
VERSION AW410870.1 GI:6936411
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 621)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Plate: LLCM56 row: P column: 10
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2961657"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 171 a 139 c 174 g 137 t
ORIGIN

Query Match 43.1%; Score 578; DB 9; Length 621;
Best Local Similarity 100.0%; Pred. No. 8.7e-147;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    764 GGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGA 823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 GGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGA 562

Qy    824 CAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAG 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 CAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAG 502

Qy    884 GGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAAT 943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 GGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAAT 442

Qy    944 GCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACT 1003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 GCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACT 382

Qy   1004 GCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGG 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 GCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGG 322

Qy   1064 CAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCT 1123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 CAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCT 262

Qy   1124 CCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAA 1183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAA 202

Qy   1184 TGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCT 1243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TGTTGTGAGGGGTTTATTTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCT 142

Qy   1244 TCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGG 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGG 82

Qy   1304 CGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1341
          |||||||||||||||||||||||||||||||||||||
Db     81 CGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 44
```

Search completed: June 6, 2002, 06:54:26
Job time: 2689 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:37 ; Search time 333.99 Seconds
(without alignments)
6893.566 Million cell updates/sec

Title: US-08-881-509-1
Perfect score: 1341
Sequence: 1 ATGAGGCAAGTGGCGAGAGT.....GCCGATGCCTTCATTAAAAT 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1341	100.0	1341	19	AAV18705	cDNA for T-cell re
2	977.4	72.9	1520	22	AAS22830	Human cDNA encodin
3	975.6	72.8	1464	21	AAZ50897	Human receptor-ass
4	972.6	72.5	1386	24	AAS62597	cDNA sequence #384
c 5	534	39.8	4658	22	AAK87082	Human immune/haema
6	513.8	38.3	1101	7	AAN60079	Sequence encoding
7	488.6	36.4	1093	22	AAS22594	Human cDNA encodin
8	446.4	33.3	834	22	AAS64171	Human T cell recep
9	446.4	33.3	834	22	AAH93935	P501S-specific T c
10	439.8	32.8	828	23	AAD23603	Human lung tumour-
11	334.4	24.9	336	19	AAV01420	Human T-cell recep
12	325	24.2	545	22	AAS22466	Human cDNA encodin
13	307.8	23.0	744	21	AAZ47236	HLA-A2/flu restric
14	307.8	23.0	744	21	AAZ47268	HLA-A2/flu restric
15	307.8	23.0	744	21	AAZ56659	TCR alpha chain an
16	307.8	23.0	744	21	AAZ56672	TCR alpha chain an
17	307.2	22.9	765	21	AAZ47241	HTLV-1 Tax/HLA-A2
18	307.2	22.9	765	21	AAZ47276	HLA-A2/HTLV-1 tax
19	307.2	22.9	765	21	AAZ56667	TCR alpha chain an
20	307.2	22.9	765	21	AAZ56677	TCR alpha chain an
21	300.6	22.4	747	21	AAZ47272	HLA-A2/HIV gag res
22	300.6	22.4	747	21	AAZ56663	TCR alpha chain an
23	299.6	22.3	1372	7	AAN60166	Sequence encoding
24	295.6	22.0	750	21	AAZ47239	HTLV-1 Tax/HLA-A2
25	295.6	22.0	750	21	AAZ47274	HLA-A2/HTLV-1 tax
26	295.6	22.0	750	21	AAZ56665	TCR alpha chain an
27	295.6	22.0	750	21	AAZ56675	TCR alpha chain an
28	289	21.6	289	20	AAH87164	Human single nucle
29	289	21.6	289	20	AAH87165	Human single nucle
30	289	21.6	289	20	AAH87166	Human single nucle
31	289	21.6	289	20	AAH87167	Human single nucle
32	283.8	21.2	407	22	AAS22702	Human cDNA encodin
33	273	20.4	330	22	ABA51249	Human breast cell
34	273	20.4	330	22	ABA69256	Human foetal liver
35	273	20.4	330	22	ABA36176	Probe #14642 for g
36	273	20.4	330	22	AAK17550	Human brain expres
37	273	20.4	330	22	AAK43358	Human bone marrow
38	273	20.4	330	22	AAI24134	Probe #14067 for g
39	273	20.4	330	22	AAI49428	Probe #18114 used
40	273	20.4	330	22	AAI09707	Probe #9698 used t
41	264.4	19.7	990	6	AAN50284	Sequence encoding
42	252.2	18.8	883	15	AAQ78793	TCR alpha-chain cl
43	250.6	18.7	251	19	AAX12362	Human biallelic po
44	250.6	18.7	251	19	AAX12363	Human biallelic po
45	250.6	18.7	251	19	AAX12364	Human biallelic po

ALIGNMENTS

```
RESULT   1
AAV18705
ID   AAV18705 standard; cDNA; 1341 BP.
XX
AC   AAV18705;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor alpha-chain.
XX
KW   Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..804
FT                   /*tag=  a
FT   sig_peptide     1..54
FT                   /*tag=  b
FT   mat_peptide     55..801
FT                   /*tag=  c
FT                   /product= T-cell_receptor_alpha-chain
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;    96DE-1025191.
XX
PR   24-JUN-1996;    96DE-1025191.
XX
```

```
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47588.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Pages 11-13; 30pp; German.
XX
CC   The present sequence encodes the alpha-chain of a human T-cell
CC   receptor (TCR), which can be used in the diagnosis, monitoring,
CC   prevention and therapy of a tumour disease, specifically renal
CC   cell carcinoma.
XX
SQ   Sequence 1341 BP; 331 A; 365 C; 294 G; 351 T; 0 other;

  Query Match             100.0%;  Score 1341;  DB 19;  Length 1341;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCTGAGTACTTTGAGCCTTGCTAAG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgaggcaagtggcgagagtgatcgtgttcctgaccctgagtactttgagccttgctaag 60

Qy      61 ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 accacccagcccatctccatggactcatatgaaggacaagaagtgaacataacctgtagc 120

Qy     121 CACAACAACATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cacaacaacattgctacaaatgattatatcacgtggtaccaacagtttcccagccaagga 180

Qy     181 CCACGATTTATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ccacgatttattattcaaggatacaagacaaaagttacaaacgaagtggcctccctgttt 240

Qy     241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 atccctgccgacagaaagtccagcactctgagcctgccccgggtttccctgagcgacact 300

Qy     301 GCTGTGTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gctgtgtactactgcctcgtgggtggttctgcaaggcaactgacctttggatctgggaca 360

Qy     361 CAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGAC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 caattgactgttttacctgatatccagaaccctgaccctgccgtgtaccagctgagagac 420

Qy     421 TCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tctaaatccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtg 480

Qy     481 TCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCT 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tcacaaagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtct 540

Qy     541 ATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 atggacttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgca 600

Qy     601 AACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 aacgccttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 660

Qy     661 TGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 tgtgatgtcaagctggtcgagaaaagctttgaaacagatacgaacctaaactttcaaaac 720

Qy     721 CTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ctgtcagtgattgggttccgaatcctcctcctgaaagtggccgggtttaatctgctcatg 780

Qy     781 ACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 acgctgcggctgtggtccagctgagatctgcaagattgtaagacagcctgtgctccctcg 840

Qy     841 CTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ctccttcctctgcattgccctcttctccctctccaaacagagggaactctcctacccc 900

Qy     901 AAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 aaggaggtgaaagctgctaccacctctgtgcccccccggcaatgccaccaactggatcct 960

Qy     961 ACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 acccgaatttatgattaagattgctgaagagctgccaaacactgctgccacccccctctgt 1020

Qy    1021 TCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 tcccttattgctgcttgtcactgcctgacattcacggcagaggcaaggctgctgcagcct 1080

Qy    1081 CCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 cccctggctgtgcacattccctcctgctccccagagactgcctccgccatcccacagatg 1140

Qy    1141 ATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTAT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 atggatcttcagtgggttctcttgggctctaggtcctggagaatgttgtgaggggtttat 1200

Qy    1201 TTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ttttttttaatagtgttcataaagaaatacatagtattcttcttctcaagacgtgggggg 1260

Qy    1261 AAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 aaattatctcattatcgaggccctgctatgctgtgtgtctgggcgtgttgtatgtcctgc 1320

Qy    1321 TGCCGATGCCTTCATTAAAAT 1341
             |||||||||||||||||||||
Db    1321 tgccgatgccttcattaaaat 1341


RESULT   2
AAS22830
ID   AAS22830 standard; cDNA; 1520 BP.
XX
AC   AAS22830;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #396.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
```

XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
DR P-PSDB; AAU14525.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Claim 1; Page 748-749; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.
XX
SQ Sequence 1520 BP; 380 A; 393 C; 351 G; 396 T; 0 other;

Query Match 72.9%; Score 977.4; DB 22; Length 1520;
Best Local Similarity 87.5%; Pred. No. 8.1e-287;
Matches 1107; Conservative 0; Mismatches 146; Indels 12; Gaps 3;

```
Qy     89 ATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTACAAATGATTATA 148
Db    233 aggaggcagagaccgtgaccctgagctgcacatatgacaccagtgagagtgattattatt 292

Qy    149 TCACGTGGTACCAACAGTTTCCCAGC-----CAAGGACCACGATTTATTATTCAAGGATA 203
Db    293 tattctggtacaagcagcctcccagcaggcagatgattctcgttattcgccaagaagctt 352

Qy    204 CAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAG----AAAGT 259
Db    353 ataagcaacagaatgcaacagagaatcgtttctctgtgaacttccagaaagcagccaaat 412

Qy    260 CCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTG---CC 316
Db    413 ccttcagtctcaagatctcagactcacagctgggggatgccgcgatgtatttctgtgctt 472

Qy    317 TCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTAC 376
Db    473 ataggagcggaagagatgacaagatcatctttggaaaagggacacgacttcatattctcc 532

Qy    377 CTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACA 436
Db    533 ccaatatccagaaccctgaccctgccgtgtaccagctgagagactctaaatccagtgaca 592

Qy    437 AGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATT 496
Db    593 agtctgtctgcctattcaccgattttgattctcaaacaaatgtgtcacaaagtaaggatt 652

Qy    497 CTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 556
Db    653 ctgatgtgtatatcacagacaaaactgtgctagacatgaggtctatggacttcaagagca 712

Qy    557 ACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACA 616
Db    713 acagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgccttcaacaaca 772

Qy    617 GCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGG 676
Db    773 gcattattccagaagacaccttcttccccagcccagaaagttcctgtgatgtcaagctgg 832

Qy    677 TCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGT 736
Db    833 tcgagaaaagctttgaaacagatacgaacctaaactttcaaaacctgtcagtgattgggt 892

Qy    737 TCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGT 796
Db    893 tccgaatcctcctcctgaaagtggccgggtttaatctgctcatgacgctgcggctgtggt 952

Qy    797 CCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATT 856
Db    953 ccagctgagatctgcaagattgtaagacagcctgtgctccctcgctccttcctctgcatt 1012

Qy    857 GCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTG 916
Db   1013 gcccctcttctccctctccaaacagagggaactctcctaccccaaggaggtgaaagctg 1072

Qy    917 CTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTATGATT 976
Db   1073 ctaccacctctgtgccccccggcaatgccaccaactggatcctacccgaatttatgatt 1132

Qy    977 AAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTT 1036
Db   1133 aagattgctgaagagctgccaaacactgctgccacccccctctgttcccttattgctgctt 1192

Qy   1037 GTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACA 1096
Db   1193 gtcactgcctgacattcacggcagaggcaaggctgctgcagcctcccctggctgtgcaca 1252

Qy   1097 TTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGG 1156
Db   1253 ttccctcctgctccccagagactgcctccgccatcccacagatgatggatcttcagtggg 1312

Qy   1157 TTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGT 1216
Db   1313 ttctcttgggctctaggtcctggagaatgttgtgaggggtttatttttttttaatagtgt 1372

Qy   1217 TCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATC 1276
Db   1373 tcataaagaaatacatagtattcttcttctcaagacgtgggggaaattatctcattatc 1432

Qy   1277 GAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATT 1336
Db   1433 gaggccctgctatgctgtgtgtctgggcgtgttgtatgtcctgctgccgatgccttcatt 1492

Qy   1337 AAAAT 1341
Db   1493 aaaat 1497
```

RESULT 3
AAZ50897
ID AAZ50897 standard; cDNA; 1464 BP.
XX
AC AAZ50897;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human receptor-associated protein cDNA from Incyte clone 1361202.
XX
KW Human receptor-associated protein; HRAP; Incyte clone 1361202;
KW cytostatic; immunomodulatory; antiinflammatory; cardiant; antianaemic;
KW antiarteriosclerotic; hepatotropic; antiarthritic antirheumatic;

```
KW   antiasthmatic; osteopathic; antiallergic; antidiabetic; dermatological;
KW   neuroprotective; diagnosis; treatment; prevention; reproductive disorder;
KW   cardiovascular; cell proliferative; autoimmune; inflammatory; allergy;
KW   gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;
KW   arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;
KW   multiple sclerosis; irritable bowel syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             93..902
FT                   /*tag=  a
FT                   /product= "HRAP"
FT   sig_peptide     93..149
FT                   /*tag=  b
FT   mat_peptide     150..899
FT                   /*tag=  c
FT                   /product= "Mature HRAP"
FT   misc_binding    921..956
FT                   /*tag=  d
FT                   /bound_moiety= "Hybridisation probe"
FT
XX
PN   WO200008155-A2.
XX
PD   17-FEB-2000.
XX
PF   06-AUG-1999;   99WO-US17777.
XX
PR   07-AUG-1998;   98US-0160065.
PR   01-SEP-1998;   98US-0098703.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Yue H,  Lal P,  Tang YT,  Gorgone GA,  Guegler KJ;
PI   Corley NC,  Baughn MR;
XX
DR   WPI; 2000-205710/18.
DR   P-PSDB; AAY69995.
XX
PT   New human receptor-associated proteins (HRAP) useful for the diagnosis,
PT   treatment and prevention of cell proliferative, autoimmune,
PT   inflammatory, reproductive, cardiovascular, and gastrointestinal
PT   disorders
XX
PS   Claim 9; Page 95; 99pp; English.
XX
CC   The present sequence is a cDNA encoding human receptor-associated protein
CC   (HRAP) from Incyte clone 1361202 obtained from LUNGNOT12 cDNA library.
CC   This sequence is expressed in haematopoietic/immune, gastrointestinal
CC   and reproductive tissues. HRAP has cytostatic, immunomodulatory,
CC   antiinflammatory, cardiant, antiarteriosclerotic, hepatotropic,
CC   antiarthritic, antirheumatic, osteopathic, antiallergic, antianaemic,
CC   antiasthmatic, antidiabetic, dermatological and neuroprotective
CC   activities. The present sequence is useful in the diagnosis, treatment
CC   and prevention of disorders associated with HRAP expression, especially
CC   cell proliferative, autoimmune/inflammatory, reproductive,
CC   cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
CC   cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
CC   asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
CC   irritable bowel syndrome).
XX
SQ   Sequence 1464 BP; 374 A; 375 C; 333 G; 382 T; 0 other;

  Query Match             72.8%;  Score 975.6;  DB 21;  Length 1464;
  Best Local Similarity   96.7%;  Pred. No. 2.8e-286;
  Matches  996;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    312 CTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGT 371
          || ||||   | ||     |  |   |  ||  ||||||| | |||||   | |||| ||
Db  * 410 ctacctctgtgctgtgagggacaatgacatgcgctttggagcagggaccagactgacagt 469

Qy    372 TTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAG 431
          •

Db    470 aaaacccaatatccagaaccctgaccctgccgtgtaccagctgagagactctaaatccag 529

Qy    432 TGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAA 491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 tgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtgtcacaaagtaa 589

Qy    492 GGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAA 551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 ggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtctatggacttcaa 649

Qy    552 GAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAA 611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 gagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgccttcaa 709

Qy    612 CAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAA 671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 caacagcattattccagaagacaccttcttccccagcccagaaagttcctgtgatgtcaa 769

Qy    672 GCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGAT 731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    770 gctggtcgagaaaagctttgaaacagatacgaacctaaactttcaaaacctgtcagtgat 829

Qy    732 TGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCT 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 tgggttccgaatcctcctcctgaaagtggccgggtttaatctgctcatgacgctgcggct 889

Qy    792 GTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCT 851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 gtggtccagctgagatctgcaagattgtaagacagcctgtgctccctcgctccttcctct 949

Qy    852 GCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAA 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 gcattgccctcttctccctctccaaacagagggaactctcccacccccaaggaggtgaa 1009

Qy    912 AGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACCCGAATTTA 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 agctgctaccacctctgtgccccccggcaatgccaccaactggatcctacccgaattta 1069

Qy    972 TGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGC 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 tgattaagattgctgaagagctgccaaacactgctgccacccccctctgttcccttattgc 1129

Qy   1032 TGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGT 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 tgcttgtcactgcctgacattcacggcagaggcaaggctgctgcagcctccccctggctgt 1189

Qy   1092 GCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCA 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 gcacattccctcctgctccccagagactgcctccgccatcccacagatgatggatcttca 1249

Qy   1152 GTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAAT 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 gtgggttctcttgggctctaggtcctgcagaatgttgtgaggggtttatttttttttaat 1309

Qy   1212 AGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCA 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 agtgttcataaagaaatacatagtattcttcttctcaagacgtgggggaaattatctca 1369

Qy   1272 TTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCT 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 ttatcgaggccctgctatgctgtgtatctgggcgtgttgtatgtcctgctgccgatgcct 1429

Qy   1332 TCATTAAAAT 1341
          ||||||||||
Db   1430 tcattaaaat 1439


RESULT   4
AAS62597
ID   AAS62597 standard; cDNA; 1386 BP.
XX
AC   AAS62597;
```

XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #384 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
XX
OS Homo sapiens.
XX
PN WO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10485.
XX
PR 06-APR-2000; 2000US-195604P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-010900/01.
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease -
XX
PS Claim 1; Page 277; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 1386 BP; 340 A; 363 C; 320 G; 363 T; 0 other;

Query Match 72.5%; Score 972.6; DB 24; Length 1386;
Best Local Similarity 98.1%; Pred. No. 2.2e-285;
Matches 984; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```
Qy    339 ACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACCC 398
          ||| | |||||    ||||  ||| ||   ||  ||||  ||||||||||||||||||||
Db    371 actcgcttttgggaaggggaaccaagtggtggtcataccaaatatccagaaccctgaccc 430

Qy    399 TGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGA 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 tgccgtgtaccagctgagagactctaaatccagtgacaagtctgtctgcctattcaccga 490

Qy    459 TTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAA 518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 ttttgattctcaaacaaatgtgtcacaaagtaaggattctgatgtgtatatcacagacaa 550

Qy    519 AACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAA 578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 aactgtgctagacatgaggtctatggacttcaagagcaacagtgctgtggcctggagcaa 610

Qy    579 CAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTT 638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 caaatctgactttgcatgtgcaaacgccttcaacaacagcattattccagaagacacctt 670

Qy    639 CTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGA 698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 cttccccagcccagaaagttcctgtgatgtcaagctggtcgagaaaagctttgaaacaga 730

Qy    699 TACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGT 758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 tacgaacctaaactttcaaaacctgtcagtgattgggttccgaatcctcctcctgaaagt 790

Qy    759 GGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTG 818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ggccgggtttaatctgctcatgacgctgcggctgtggtccagctgagatctgcaagattg 850

Qy    819 TAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAA 878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 taagacagcctgtgctccctcgctccttcctctgcattgccccttcttctccctctccaaa 910

Qy    879 CAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCG 938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 cagagggaactctcctacccccaaggaggtgaaagctgctaccacctctgtgccccccg 970

Qy    939 GCAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAA 998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 gcaatgccaccaactggatcctacccgaatttatgattaagattgctgaagagctgccaa 1030

Qy    999 ACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGC 1058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 acactgctgccacccccctctgttcccttattgctgcttgtcactgcctgacattcacggc 1090

Qy   1059 AGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGAC 1118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 agaggcaaggctgctgcagcctccccctggctgtgcacattccctcctgctccccagagac 1150

Qy   1119 TGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTG 1178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 tgcctccgccatcccacagatgatggatcttcagtgggttctcttgggctctaggtcctg 1210

Qy   1179 GAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATT 1238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 cagaatgttgtgaggggtttatttttttttaatagtgttcataaagaaatacatagtatt 1270

Qy   1239 CTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGT 1298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 cttcttctcaagacgtgggggaaattatctcattatcgaggccctgctatgctgtgtat 1330

Qy   1299 CTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAAT 1341
          |||||||||||||||||||||||||||||||||||||||||||
Db   1331 ctgggcgtgttgtatgtcctgctgccgatgccttcattaaaat 1373
```

RESULT 5
AAK87082/c
ID AAK87082 standard; DNA; 4658 BP.
XX
AC AAK87082;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41894.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

```
XX
PS   Disclosure; SEQ ID NO 41894; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 4658 BP; 1155 A; 1156 C; 1182 G; 1165 T; 0 other;

  Query Match             39.8%;  Score 534;  DB 22;  Length 4658;
  Best Local Similarity   97.3%;  Pred. No. 1.6e-151;
  Matches  543;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy     784 CTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 843
Db     576 CTTCTGTTCTTCCTCATTCCAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTC 517

Qy     844 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 903
Db     516 CTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAG 457

Qy     904 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTGGATCCTACC 963
Db     456 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAATGCCACCAACTGGATCCTACA 397

Qy     964 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 1023
Db     396 CGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGTTCC 337

Qy    1024 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 1083
Db     336 CTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCC 277

Qy    1084 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 1143
Db     276 CTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG 217

Qy    1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTT 1203
Db     216 GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCCGGAGAATGTTGTGAGGGGTTTATTTT 157

Qy    1204 TTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 1263
Db     156 TTTTTAATAGTGTTCATAAAGAAAGACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAA 97

Qy    1264 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 1323
Db      96 TTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATGTCCTGCTGC 37

Qy    1324 CGATGCCTTCATTAAAAT 1341
Db      36 CGATGCCTTCATTAAAAT 19

RESULT   6
AAN60079
ID   AAN60079 standard; cDNA; 1101 BP.
XX
AC   AAN60079;
XX
DT   23-JUL-1991  (first entry)
XX
DE   Sequence encoding a polypeptide which is at least part of the alpha
DE   chain of T cell antigen receptor.
XX
KW   Tumour cell; diagnosis; T cell antigen receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             190..249
FT                   /*tag= a
FT                   /product= leader sequence
FT   CDS             250..526
FT                   /*tag= b
FT                   /product= variable region
FT   CDS             527..542
FT                   /*tag= c
FT                   /product= diversity region
FT   CDS             543..594
FT                   /*tag= d
FT                   /product= joining region
FT   CDS             595..945
FT                   /*tag= e
FT                   /product= constant region
FT   CDS             946..1005
FT                   /*tag= f
FT                   /product= transmembrane region
FT   CDS             1006..1023
FT                   /*tag= g
FT                   /product= cytoplasmic region
XX
PN   EP200350-A.
XX
PD   05-NOV-1986.
XX
PF   25-MAR-1986;    86EP-0021945.
XX
PR   15-APR-1985;    85US-0723306.
XX
PA   (ONTA-) ONTARIO CANCER INST.
XX
PI   Mak TW;
XX
DR   WPI; 1986-292977/45.
DR   P-PSDB; AAP60065.
XX
PT   Nucleic acid encoding T-cell antigen receptor polypeptide - used in
PT   identifying T-cells etc.
XX
PS   Claim 4; Fig 4; 23pp; English.
XX
CC   AAN60079 was generated from the Jurkat human leukaemia T cell line. It
CC   is contained in T cell clone pY14. The labelled nucleic acid and
CC   monoclonal or polyclonal antibodies to the polypeptides may be used
CC   to determine whether unknown cells, e.g. tumour cells, are T cells.
XX
SQ   Sequence 1101 BP; 263 A; 314 C; 256 G; 267 T; 1 other;

  Query Match             38.3%;  Score 513.8;  DB 7;  Length 1101;
  Best Local Similarity   96.7%;  Pred. No. 9.7e-146;
  Matches  524;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy     341 TGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTG 400
Db     560 taatctttggatcagggaccagactcagcatccggccaaatatccagaaccctgaccctg 619

Qy     401 CCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATT 460
Db     620 ccgtgtaccagctgagagactctaaatccagtgacaagtctgtctgcctattcaccgatt 679
```

```
Qy     461 TTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAA 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     680 ttgattctcaaacaaatgtgtcacaaagtaaggattctgatgtgtatatcacagacaaaa 739

Qy     521 CTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACA 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     740 ctgtgctagacatgaggtctatggacttcaagagcaacagtgctgtggcctggagcaaca 799

Qy     581 AATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCT 640
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     800 aatctgactttgcatgngcaaacgccttcaacaacagcattattccagaagacaccttct 859

Qy     641 TCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATA 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     860 tccccagcccagaaagttcctgtgatgtcaagctggtcgagaaaagctttgaaacagata 919

Qy     701 CGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGG 760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     920 cgaacctaaactttcaaaacctgtcagtgattgggttccgaatcctcctcctgaaagtgg 979

Qy     761 CCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTA 820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     980 ccgggtttaatctgctcatgacgctgcggctgtggtccagctgagatctgcaagattgta 1039

Qy     821 AGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACA 880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1040 agacagcctgtgctccctcgctccttcctctgcattgccccctcttctccctctccaaaca 1099

Qy     881 GA 882
           ||
Db    1100 ga 1101
```

```
RESULT   7
AAS22594
ID   AAS22594 standard; cDNA; 1093 BP.
XX
AC   AAS22594;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #160.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14289.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 390-391; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity.  The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy.  Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 1093 BP; 287 A; 278 C; 241 G; 283 T; 4 other;
```

Query Match 36.4%; Score 488.6; DB 22; Length 1093;
Best Local Similarity 93.6%; Pred. No. 4.6e-138;
Matches 542; Conservative 0; Mismatches 34; Indels 3; Gaps 3;

```
Qy     322 GGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGAT 381
           || ||   |||      ||| |||||||||    ||||| ||  | |   |    ||  ||
Db     376 ggaggaggtgctgacggactcacctttggcaaagggactcatctaatcatccagccctat 435

Qy     382 ATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCT 441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 atccagaaccctgaccctgccgtgtaccagctgagagactctaaatccagtgacaagtct 495

Qy     442 GTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGAT 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 gtctgcctattcaccgattttgattctcaaacaaatgtgtcacaaagtaaggattctgat 555

Qy     502 GTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGT 561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     556 gtgtatatcacagacaaaactgtgctagacatgaggtctatggacttcaagagcaacagt 615

Qy     562 GCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATT 621
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     616 gctgtggcctggagcaacaaatctgactttgcatgtgcaaacgccttcaacaacagcatt 675

Qy     622 ATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAG 681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     676 attccagaagacaccttcttccccagcccagaaagttcctgtgatgtcaagctggtcgag 735

Qy     682 AAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGA 741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     736 aaaagctttgaaacagatacgaacctaaactttcaaaacctgtcagtgattgggttccga 795

Qy     742 ATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGT-GGTCCAG 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||
Db     796 atcctcctcctgaaagtggccgggtttaatctgctcatgacgctgcggctgtgggtccag 855

Qy     801 CTGAGATCTGCAAGATTGTAA-GACAGCCTGTGCTCCCTCGCTCCTTCCTCT-GCATTGC 858
           |||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||| |||
Db     856 ctgagatctgcaagattgtaaggacagcctgtgctccctcgctccttcctctggcattgc 915

Qy     859 CCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACC 897
           ||||||||||||||||||||||||||||    || || ||
Db     916 ccctcttctccctctccaaacagaggggaacttcttcc 954
```

RESULT 8
AAS64171
ID AAS64171 standard; cDNA; 834 BP.
XX
AC AAS64171;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human T cell receptor Va chain 4E5 for prostate protein P501S cDNA.
XX
KW Rabbit; prostate cancer; ss; cytostatic; immunostimulant; tumour;
KW antibody.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR P-PSDB; AAU69943.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Example 25; Page 560; 579pp; English.
XX
CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence encodes an antibody
CC molecule raised against a prostate specific polypeptide of the invention.
XX
SQ Sequence 834 BP; 239 A; 199 C; 194 G; 201 T; 1 other;

Query Match 33.3%; Score 446.4; DB 22; Length 834;
Best Local Similarity 77.3%; Pred. No. 2.8e-125;
Matches 575; Conservative 0; Mismatches 151; Indels 18; Gaps 2;

```
Qy      79 ATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTACA 138
           |||  |    | |||    | |  ||||   |   ||| |   |  ||| || || |  |
Db      91 atgttcgtgcaggaaaaggaggctgtgactctggactgcacatatgacaccagtgatcaa 150

Qy     139 AATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTCA- 197
           | | ||  | ||   |||||| | |||     |||    | |      ||| ||||| |
Db     151 agttatggtctcttctggtacaagcagcccagcagtggggaaatgatttttcttatttat 210

Qy     198 --------AGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
                     || ||  ||| |  ||| | |        |||  ||  | |||
Db     211 caggggtcttatgacgagcaaaatgcaacagaaggtcgctactcattgaatttccagaag 270

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
           |  ||||| |||  ||  ||   | |  ||          ||| | ||| | ||  ||||
Db     271 gcaagaaaatccgccaaccttgtcatctccgcttcacaactgggggactcagcaatgtat 330

Qy     310 TACTGCCTCGTG---------GGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360
           | |||     ||         |  ||    | ||   |||| ||||||||  | || ||
Db     331 ttctgtgcaatgagagagggcgcgggaggaggaaacaaactcacctttgggacaggcact 390

Qy     361 CAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGAC 420
           ||  | |  ||   ||  ||||||||||||||||||||||||||||||||||||||||
Db     391 cagctaaaagtggaactcaatatccagaaccctgaccctgccgtgtaccagctgagagac 450

Qy     421 TCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     451 tctaaatccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtg 510

Qy     481 TCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     511 tcacaaagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtct 570

Qy     541 ATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     571 atggacttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgca 630

Qy     601 AACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     631 aacgccttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 690

Qy     661 TGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     691 tgtgatgtcaagctggtcgagaaaagctttgaaacagatacgaacctaaactttcaaaac 750

Qy     721 CTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     751 ctgtcagtgattgggttccgaatcctcctcctgaaagtggccgggtttaatctgctcatg 810

Qy     781 ACGCTGCGGCTGTGGTCCAGCTGA 804
           ||||||||||||||||||||||||
Db     811 acgctgcggctgtggtccagctga 834
```

RESULT 9
AAH93935
ID AAH93935 standard; cDNA; 834 BP.
XX
AC AAH93935;
XX
DT 04-OCT-2001 (first entry)
XX
DE P501S-specific T cell clone 4E5 Va chain T cell receptor cDNA sequence.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.

XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Example 24; Page 523-524; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 834 BP; 239 A; 199 C; 194 G; 201 T; 1 other;

Query Match 33.3%; Score 446.4; DB 22; Length 834;
Best Local Similarity 77.3%; Pred. No. 2.8e-125;
Matches 575; Conservative 0; Mismatches 151; Indels 18; Gaps 2;

```
Qy      79 ATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTACA 138
           |||  |   | |||    | |  ||||   |   ||| |   |  ||| || || |  |
Db      91 atgttcgtgcaggaaaaggaggctgtgactctggactgcacatatgacaccagtgatcaa 150

Qy     139 AATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTCA- 197
           | | ||  | ||   |||||| | |||     |||    | |      ||| |||||  |
Db     151 agttatggtctcttctggtacaagcagcccagcagtggggaaatgatttttcttatttat 210

Qy     198 --------AGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
                   || ||  ||| |  ||| | |       |||  ||  | |||
Db     211 caggggtcttatgacgagcaaaatgcaacagaaggtcgctactcattgaatttccagaag 270

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
           |  ||||| |||  ||  ||   | |  ||                |||  | ||| | ||  ||||
Db     271 gcaagaaaatccgccaaccttgtcatctccgcttcacaactgggggactcagcaatgtat 330

Qy     310 TACTGCCTCGTG---------GGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360
           | |||     ||          |  ||    | ||   |||| ||||||||  | || ||
Db     331 ttctgtgcaatgagagagggcgcgggaggaggaaacaaactcacctttgggacaggcact 390

Qy     361 CAATTGACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGAC 420
           ||  | |  ||   ||  ||||||||||||||||||||||||||||||||||||||||||
Db     391 cagctaaaagtggaactcaatatccagaaccctgaccctgccgtgtaccagctgagagac 450

Qy     421 TCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     451 tctaaatccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtg 510

Qy     481 TCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     511 tcacaaagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtct 570

Qy     541 ATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     571 atggacttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgca 630

Qy     601 AACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     631 aacgccttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 690

Qy     661 TGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     691 tgtgatgtcaagctggtcgagaaaagctttgaaacagatacgaacctaaactttcaaaac 750

Qy     721 CTGTCAGTGATTGGGTTCCGAATCCTCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     751 ctgtcagtgattgggttccgaatcctcctcctgaaagtggccgggtttaatctgctcatg 810

Qy     781 ACGCTGCGGCTGTGGTCCAGCTGA 804
           ||||||||||||||||||||||||
Db     811 acgctgcggctgtggtccagctga 834
```

RESULT 10
AAD23603
ID AAD23603 standard; cDNA; 828 BP.
XX
AC AAD23603;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific T cell receptor alpha chain cDNA.
XX
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer;
KW T cell receptor alpha chain; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..828
FT /*tag= a
FT /product= "Human lung tumour-specific T cell receptor
FT alpha chain"
XX
PN WO200172295-A2.
XX
PD 04-OCT-2001.
XX
PF 28-MAR-2001; 2001WO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SR, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
DR P-PSDB; AAE13849.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
PS Claim 1; Page 376; 378pp; English.
XX
CC The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific T cell receptor alpha
CC chain.

XX
SQ Sequence 828 BP; 235 A; 208 C; 182 G; 203 T; 0 other;

Query Match 32.8%; Score 439.8; DB 23; Length 828;
Best Local Similarity 96.4%; Pred. No. 2.8e-123;
Matches 450; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy    338 AACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACC 397
          | |||  |||||||   || ||  || |||| ||    ||  ||||||||||||||||||
Db    362 agctggtctttggaaagggaacgaaactgacagtaaacccatatatccagaaccctgacc 421

Qy    398 CTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCG 457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ctgccgtgtaccagctgagagactctaaatccagtgacaagtctgtctgcctattcaccg 481

Qy    458 ATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACA 517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 attttgattctcaaacaaatgtgtcacaaagtaaggattctgatgtgtatatcacagaca 541

Qy    518 AAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCA 577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 aaactgtgctagacatgaggtctatggacttcaagagcaacagtgctgtggcctggagca 601

Qy    578 ACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCT 637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 acaaatctgactttgcatgtgcaaacgccttcaacaacagcattattccagaagacacct 661

Qy    638 TCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAG 697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 tcttccccagcccagaaagttcctgtgatgtcaagctggtcgagaaaagctttgaaacag 721

Qy    698 ATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAG 757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 atacgaacctaaactttcaaaacctgtcagtgattgggttccgaatcctcctcctgaaag 781

Qy    758 TGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGA 804
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    782 tggccgggtttaatctgctcatgacgctgcggctgtggtccagctga 828
```

RESULT 11
AAV01420
ID AAV01420 standard; cDNA to mRNA; 336 BP.
XX
AC AAV01420;
XX
DT 19-MAY-1998 (first entry)
XX
DE Human T-cell receptor alpha-chain constant region encoding cDNA.
XX
KW Human; T-cell receptor; alpha-chain constant region; antigen-specific;
KW immunosuppressant; humoral; cell mediated immune response; allergy;
KW hypersensitivity; autoimmune reaction; transplant rejection; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..336
FT /*tag= a
FT /product= "T-cell receptor alpha-chain constant region"
FT /note= "no stop codon"
XX
PN WO9743411-A1.
XX
PD 20-NOV-1997.
XX
PF 09-MAY-1997; 97WO-JP01565.
XX
PR 29-MAY-1996; 96JP-0135572.
PR 10-MAY-1996; 96JP-0116101.
XX
PA (KIRI) KIRIN BEER KK.
XX
PI Honma N, Mikayama T, Yuyama N;
XX
DR WPI; 1998-008880/01.
DR P-PSDB; AAW36112.
XX
PT Immunosuppressant peptide containing T-cell receptor alpha-chain
PT sequence - are not antigen-specific and do not induce antibody
PT production.
XX
PS Example 10; Page 45-46; 63pp; Japanese.
XX
CC The present sequence encodes human T-cell receptor alpha-chain
CC constant region. The protein is an immunosuppressant which is not
CC antigen-specific and suppresses both humoral and cell -mediated immune
CC reactions. It can be used for treatment and/or prevention of delayed
CC hypersensitivity reactions, allergies and autoimmune reactions, and
CC inhibition of transplant rejection. The protein does not induce the
CC formation of antibodies against them to any significant extent.
XX
SQ Sequence 336 BP; 105 A; 84 C; 65 G; 82 T; 0 other;

Query Match 24.9%; Score 334.4; DB 19; Length 336;
Best Local Similarity 99.7%; Pred. No. 2.1e-91;
Matches 335; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    382 ATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCT 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atccagaaccctgaccctgccgtgtaccagctgagagactctaaatccagtgacaagtct 60

Qy    442 GTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGAT 501
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     61 gtctgcctattcaccgattttgattctcaaacaaatgtctcacaaagtaaggattctgat 120

Qy    502 GTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGT 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gtgtatatcacagacaaaactgtgctagacatgaggtctatggacttcaagagcaacagt 180

Qy    562 GCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATT 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gctgtggcctggagcaacaaatctgactttgcatgtgcaaacgccttcaacaacagcatt 240

Qy    622 ATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAG 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 attccagaagacaccttcttccccagcccagaaagttcctgtgatgtcaagctggtcgag 300

Qy    682 AAAAGCTTTGAAACAGATACGAACCTAAACTTTCAA 717
          |||||||||||||||||||||||||||||||||||
Db    301 aaaagctttgaaacagatacgaacctaaactttcaa 336
```

RESULT 12
AAS22466
ID AAS22466 standard; cDNA; 545 BP.
XX
AC AAS22466;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA encoding a novel human protein #32.
XX
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.

XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
DR P-PSDB; AAU14161.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Claim 1; Page 220-221; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.
XX
SQ Sequence 545 BP; 164 A; 121 C; 111 G; 149 T; 0 other;

Query Match 24.2%; Score 325; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 2e-88;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCTGAGTACTTTGAGCCTTGCTAAG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgaggcaagtggcgagagtgatcgtgttcctgaccctgagtactttgagccttgctaag 60

Qy      61 ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 accacccagcccatctccatggactcatatgaaggacaagaagtgaacataacctgtagc 120

Qy     121 CACAACAACATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cacaacaacattgctacaaatgattatatcacgtggtaccaacagtttcccagccaagga 180

Qy     181 CCACGATTTATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ccacgatttattattcaaggatacaagacaaaagttacaaacgaagtggcctccctgttt 240

Qy     241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  *  241 atccctgccgacagaaagtccagcactctgagcctgccccgggtttccctgagcgacact 300

Qy     301 GCTGTGTACTACTGCCTCGTGGGTG 325
```

```
             |||||||||||||||||||||||||
Db     301 gctgtgtactactgcctcgtgggtg 325
```

RESULT 13
AAZ47236
ID AAZ47236 standard; DNA; 744 BP.
XX
AC AAZ47236;
XX
DT 11-APR-2000 (first entry)
XX
DE HLA-A2/flu restricted matrix TCR-alpha/c-jun fusion coding sequence.
XX
KW Fusion protein; T-cell receptor; c-jun; leucine zipper motif; tumour; ss;
KW influenza virus matrix protein epitope; major histocompatibility complex;
KW immune system.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9960119-A2.
XX
PD 25-NOV-1999.
XX
PF 19-MAY-1999; 99WO-GB01583.
XX
PR 19-MAY-1998; 98GB-0010759.
PR 29-SEP-1998; 98GB-0021129.
XX
PA (AVID-) AVIDEX LTD.
XX
PI Jakobsen BK, Boulter JM;
XX
DR WPI; 2000-072439/06.
XX
PT Synthetic multivalent receptor complexes used to detect Major
PT Histocompatibility Complex-peptide complexes, and delivering
PT therapeutic agents to target cells -
XX
PS Example 4; Fig 4; 155pp; English.
XX
CC This sequence represents the coding region for a fusion protein
CC comprising a T-cell receptor (TCR) alpha chain linked to a c-jun leucine
CC zipper motif. The TCR gene is taken from clone JM22 and is targeted to
CC an HLA-A2 influenza virus matrix protein epitope. The invention relates
CC to a synthetic multivalent complex for binding to a major
CC histocompatibility complex (MHC)-peptide complex. The multivalent TCR
CC complexes are useful for detecting MHC-peptide complexes and delivery
CC of therapeutic agents to target cells, in vivo. The specificity of the
CC TCRs enables the localization of the liposome-contained drugs to the
CC desired target site such as a tumour or virus-infected cell. This would
CC be useful in many situations and in particular against tumours because
CC not all cells in the tumour present antigens and therefore are not all
CC detected by the immune system.
XX
SQ Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;

Query Match 23.0%; Score 307.8; DB 21; Length 744;
Best Local Similarity 83.9%; Pred. No. 4.1e-83;
Matches 348; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy     246 TGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGT 305
           ||  | ||||||  |||  ||||  | |  |  ||     |   | || || |   |
Db     198 tgatgcaagaaaggacagttctctccacatcactgcggcccagcctggtgatacaggcct 257

Qy     306 GTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATT 365
           |||  |||    |  |  ||      | |  | || | |||||||   || ||  || |
Db     258 ctacctctgtgcaggagcgggaagccaaggaaatctcatctttggaaaaggcactaaact 317

Qy     366 GACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAA 425
             |||||  |||  ||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    318 ctctgttaaaccaaatatccagaaccctgaccctgccgtgtaccagctgagagactctaa 377

Qy    426 ATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACA 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 atccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtgtcaca 437

Qy    486 AAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGA 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 aagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtctatgga 497

Qy    546 CTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 cttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgc 557

Qy    606 CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 cttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 612


RESULT 14
AAZ47268
ID   AAZ47268 standard; DNA; 744 BP.
XX
AC   AAZ47268;
XX
DT   11-APR-2000  (first entry)
XX
DE   HLA-A2/flu restricted matrix TCR-alpha/c-jun fusion coding sequence.
XX
KW   Fusion protein; T-cell receptor; c-jun; leucine zipper motif; tumour; ss;
KW   influenza virus matrix protein epitope; major histocompatibility complex;
KW   immune system.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9960119-A2.
XX
PD   25-NOV-1999.
XX
PF   19-MAY-1999;   99WO-GB01583.
XX
PR   19-MAY-1998;   98GB-0010759.
PR   29-SEP-1998;   98GB-0021129.
XX
PA   (AVID-) AVIDEX LTD.
XX
PI   Jakobsen BK,  Boulter JM;
XX
DR   WPI; 2000-072439/06.
DR   P-PSDB; AAY56076.
XX
PT   Synthetic multivalent receptor complexes used to detect Major
PT   Histocompatibility Complex-peptide complexes, and delivering
PT   therapeutic agents to target cells -
XX
PS   Example 10; Fig 21; 155pp; English.
XX
CC   This sequence represents the coding region for a fusion protein
CC   comprising a T-cell receptor (TCR) alpha chain linked to a c-jun leucine
CC   zipper motif. The TCR gene is taken from clone JM22 and is targeted to
CC   an HLA-A2 influenza virus matrix protein epitope. The sequence also
CC   contains mutations in the 5' end of the gene to enhance expression of
CC   the protein in E. coli. The invention relates to a synthetic multivalent
CC   complex for binding to a major histocompatibility complex (MHC)-peptide
CC   complex. The multivalent TCR complexes are useful for detecting
CC   MHC-peptide complexes and delivery of therapeutic agents to target cells,
CC   in vivo. The specificity of the TCRs enables the localization of the
CC   liposome-contained drugs to the desired target site such as a tumour or
CC   virus-infected cell. This would be useful in many situations and in
CC   particular against tumours because not all cells in the tumour present
CC   antigens and therefore are not all detected by the immune system.
XX

SQ   Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;

  Query Match             23.0%;  Score 307.8;  DB 21;  Length 744;
  Best Local Similarity   83.9%;  Pred. No. 4.1e-83;
  Matches  348;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    246 TGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGT 305
          ||  |  ||||||  |||  ||||   | |  |   ||      |   | || || |   |
Db    198 tgatgcaagaaaggacagttctctccacatcactgcggcccagcctggtgatacaggcct 257

Qy    306 GTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATT 365
          |||  |||    |  |  ||      | |  | || | |||||||   || ||  || |
Db    258 ctacctctgtgcaggagcgggaagccaaggaaatctcatctttggaaaaggcactaaact 317

Qy    366 GACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAA 425
            |||||  |||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    318 ctctgttaaaccaaatatccagaaccctgaccctgccgtgtaccagctgagagactctaa 377

Qy    426 ATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACA 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 atccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtgtcaca 437

Qy    486 AAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGA 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 aagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtctatgga 497

Qy    546 CTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 cttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgc 557

Qy    606 CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 cttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 612


RESULT 15
AAZ56659
ID   AAZ56659 standard; DNA; 744 BP.
XX
AC   AAZ56659;
XX
DT   22-MAR-2000  (first entry)
XX
DE   TCR alpha chain and c-jun fusion protein encoding DNA - Fig 10.
XX
KW   T cell receptor; TCR; soluble; cytotoxic; immunostimulatory; MHC;
KW   major histocompatibility complex; vaccine; infection; detection;
KW   diagnosis; autoimmune disease; tumour; viral disease; cancer;
KW   metastasis; graft rejection; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9960120-A2.
XX
PD   25-NOV-1999.
XX
PF   19-MAY-1999;   99WO-GB01588.
XX
PR   19-MAY-1998;   98GB-0010759.
PR   29-SEP-1998;   98GB-0021129.
XX
PA   (AVID-) AVIDEX LTD.
XX
PI   Jakobsen BK,  Bell JI,  Gao GF,  Willcox BE,  Boulter JM;
XX
DR   WPI; 2000-086595/07.
DR   P-PSDB; AAY57852.
XX
PT   Refolded recombinant T-cell receptors used to detect the efficacy of T
PT   cell vaccines -
XX
```

PS Example 7; Fig 10; 125pp; English.
XX
CC The present invention describes a refolded recombinant T-cell receptor
CC (TCR) comprising: (a) a recombinant TCR alpha or gamma chain
CC extracellular domain having a first heterologous C-terminal dimerisation
CC peptide; and (b) a recombinant TCR beta or delta chain extracellular
CC domain having a second C-terminal dimerisation peptide which is
CC specifically heterodimerised with the first dimerisation peptide to form
CC a heterodimerisation domain. The recombinant TCR can be used for
CC investigating specific TCR-MHC (major histocompatibility complex)
CC interactions, and also as a diagnostic tool to detect infection or
CC autoimmune disease markers, or to detect the efficacy of T cell vaccines.
CC The TCR could be used to deliver therapeutic agents to cells presenting a
CC particular antigen, e.g. delivering a toxin to a tumour to eradicate it.
CC The TCRs would be beneficial in treating viral diseases for which drugs
CC exist, by allowing for the drug to be released in the near vicinity of
CC infected cells. For cancer, the localisation in the vicinity of tumours
CC or metastasis using the TCRs would enhance the effect of toxin or
CC immunostimulants. In autoimmune diseases, using the TCRs,
CC immunosuppressive drugs could be released slowly, having more local
CC effect over a longer time-span while minimally affecting the overall
CC immuno-capacity of the subject. In the prevention of graft rejection,
CC the effect of immunosuppressive drugs could be optimised in the same
CC way using the TCRs. The present sequence encodes a TCR fusion protein
CC from an example from the present invention.
XX
SQ Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;

Query Match 23.0%; Score 307.8; DB 21; Length 744;
Best Local Similarity 83.9%; Pred. No. 4.1e-83;
Matches 348; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy    246 TGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGT 305
          ||  |  ||||||  |||  ||||   | |  |   ||      |   | || || |   |
Db    198 tgatgcaagaaaggacagttctctccacatcactgcggcccagcctggtgatacaggcct 257

Qy    306 GTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATT 365
           |||  |||    |  |  ||      | |  | || | |||||||   || ||  || |
Db    258 ctacctctgtgcaggagcgggaagccaaggaaatctcatctttggaaaaggcactaaact 317

Qy    366 GACTGTTTTACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAA 425
            |||||  |||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    318 ctctgttaaaccaaatatccagaaccctgaccctgccgtgtaccagctgagagactctaa 377

Qy    426 ATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACA 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 atccagtgacaagtctgtctgcctattcaccgattttgattctcaaacaaatgtgtcaca 437

Qy    486 AAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGA 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 aagtaaggattctgatgtgtatatcacagacaaaactgtgctagacatgaggtctatgga 497

Qy    546 CTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGC 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 cttcaagagcaacagtgctgtggcctggagcaacaaatctgactttgcatgtgcaaacgc 557

Qy    606 CTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 cttcaacaacagcattattccagaagacaccttcttccccagcccagaaagttcc 612
```

Search completed: June 6, 2002, 07:00:20
Job time: 3043 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:38 ; Search time 78.65 Seconds
(without alignments)
4188.106 Million cell updates/sec

Title: US-08-881-509-1
Perfect score: 1341
Sequence: 1 ATGAGGCAAGTGGCGAGAGT.....GCCGATGCCTTCATTAAAAT 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	448	33.4	822	1	US-07-936-267A-8	Sequence 8, Appli
2	299.6	22.3	1372	6	5189147-2	Patent No. 5189147
3	264.4	19.7	990	4	US-09-082-593-9	Sequence 9, Appli
4	247.4	18.4	804	1	US-08-416-336-1	Sequence 1, Appli
5	57	4.3	747	4	US-09-140-084-25	Sequence 25, Appl
6	55.8	4.2	7218	1	US-08-232-463-14	Sequence 14, Appl
7	41.8	3.1	93	2	US-08-411-098-6	Sequence 6, Appli
8	39.8	3.0	645	2	US-08-480-753-7	Sequence 7, Appli
9	37.2	2.8	4691	4	US-08-591-632-43	Sequence 43, Appl
10	37.2	2.8	6166	4	US-08-591-632-51	Sequence 51, Appl
11	36.8	2.7	330	4	US-09-025-769B-46	Sequence 46, Appl
12	36.4	2.7	7655	1	US-08-619-554-1	Sequence 1, Appli
13	35.6	2.7	402	1	US-08-405-034-3	Sequence 3, Appli
14	35.6	2.7	646	1	US-08-300-386A-2	Sequence 2, Appli
15	35.6	2.7	646	3	US-08-931-645-2	Sequence 2, Appli
16	35.6	2.7	646	5	PCT-US94-01258-2	Sequence 2, Appli
17	35.6	2.7	646	5	PCT-US95-11235-2	Sequence 2, Appli
18	35.6	2.7	714	1	US-07-971-160-1	Sequence 1, Appli
19	35.6	2.7	714	1	US-08-336-241-1	Sequence 1, Appli
20	35.6	2.7	714	2	US-08-465-273-1	Sequence 1, Appli
21	35.6	2.7	714	2	US-09-119-024-1	Sequence 1, Appli
22	35.6	2.7	714	2	US-08-417-226-1	Sequence 1, Appli
23	35.6	2.7	714	4	US-09-196-131-1	Sequence 1, Appli
24	35.6	2.7	754	1	US-07-971-160-5	Sequence 5, Appli
25	35.6	2.7	754	1	US-08-336-241-5	Sequence 5, Appli
26	35.6	2.7	754	2	US-08-465-273-5	Sequence 5, Appli
27	35.6	2.7	754	2	US-09-119-024-5	Sequence 5, Appli
28	35.6	2.7	754	2	US-08-417-226-5	Sequence 5, Appli
29	35.6	2.7	754	4	US-09-196-131-5	Sequence 5, Appli
30	35.6	2.7	1738	1	US-07-971-160-9	Sequence 9, Appli
31	35.6	2.7	1738	1	US-08-336-241-9	Sequence 9, Appli
32	35.6	2.7	1738	2	US-08-465-273-9	Sequence 9, Appli
33	35.6	2.7	1738	2	US-09-119-024-9	Sequence 9, Appli
34	35.6	2.7	1738	2	US-08-417-226-9	Sequence 9, Appli
35	35.6	2.7	1738	4	US-09-196-131-9	Sequence 9, Appli
36	35	2.6	396	2	US-08-345-321-3	Sequence 3, Appli
c 37	34.8	2.6	771	4	US-08-991-789A-241	Sequence 241, App
c 38	34.8	2.6	771	4	US-09-062-451-241	Sequence 241, App
c 39	34.8	2.6	1770	3	US-09-006-636-1	Sequence 1, Appli
c 40	34.8	2.6	1770	4	US-09-006-632-1	Sequence 1, Appli
c 41	34.8	2.6	6000	3	US-09-006-636-9	Sequence 9, Appli
c 42	34.8	2.6	6000	4	US-09-006-632-9	Sequence 9, Appli
c 43	34.4	2.6	289	4	US-09-007-005-17	Sequence 17, Appl
c 44	34.4	2.6	289	4	US-09-244-796-17	Sequence 17, Appl
45	34.2	2.6	312	3	US-08-767-128-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-07-936-267A-8
; Sequence 8, Application US/07936267A
; Patent No. 5445940
; GENERAL INFORMATION:
;   APPLICANT: Michael B. Brenner and Harout DerSimonian
;   TITLE OF INVENTION: V'12.1 T-Cell Receptor-
;   TITLE OF INVENTION: Specific Reagents For Diagnosing And Treating
;   TITLE OF INVENTION: Rheumatoid Arthritis
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Dana-Farber Cancer Institute, Inc.
;     STREET: 44 Binney Street
;     CITY: Boston
;     STATE: Massachusetts
;     COUNTRY: USA
;     ZIP: 02115
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.50 inch, 720Kb
;     MEDIUM TYPE: storage
;     COMPUTER: IBM Personal System 2; Model 30
;     OPERATING SYSTEM: MS/DOS
;     SOFTWARE: WordPerfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/936,267A
;     FILING DATE: 26-AUG-1992
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/750,986
;     FILING DATE: 28-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: HART, JULIA D.
;     REGISTRATION NUMBER: 33132
;     REFERENCE/DOCKET NUMBER: DFCI-121.1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617) 632-4016
;     TELEFAX: (617) 632-4012
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 822 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: no
;   ANTI-SENSE: no
;   ORIGINAL SOURCE:
;     ORGANISM: Homo sapien
;     DEVELOPMENTAL STAGE: adult
;     CELL TYPE: Human T-cell leukemia line HBM-MLT
```

```
;    FEATURE:
;      NAME/KEY:  T cell receptor \'-chain-encoding cDNA sequence
;      LOCATION:  1-822
;      OTHER INFORMATION:  Encoded amino acids -20 to -1 are
;      OTHER INFORMATION:  predicted to be the leader sequence; encoded amino acids 1-95
;      OTHER INFORMATION:  predicted to be the variable region; encoded amino acids 96-1
;      OTHER INFORMATION:  predicted to be the joining region; encoded amino acids 112-2
;      OTHER INFORMATION:  predicted to be the constant region; encoded amino acids 229-
;      OTHER INFORMATION:  are predicted to be the transmembrane region; and encoded ami
;      OTHER INFORMATION:  acids 249-253 are predicted to be the cytoplasmic region.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Sim et al.
;      TITLE:  Primary structure of human T-cell receptor \'-chain
;      JOURNAL:  Nature
;      VOLUME:  312
;      PAGES:  771-775
;      DATE:  1984
US-07-936-267A-8
```

Query Match 33.4%; Score 448; DB 1; Length 822;
Best Local Similarity 77.8%; Pred. No. 4.3e-123;
Matches 557; Conservative 0; Mismatches 150; Indels 9; Gaps 1;

```
Qy      98 AAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTACAAATGATTATATCACGTGGT 157
           | || |||| | |   ||||    |  | | |  || ||| | | ||||  |    ||||
Db     107 AGGATGTGACCTTGGACTGTGTGTATGAAACCCGTGATACTACTTATTACTTATTCTGGT 166

Qy     158 ACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTCAAGGATAC---------AAGA 208
           || | ||    || ||   || |      ||  ||||||   |  ||          ||
Db     167 ACAAGCAACCACCAAGTGGAGAATTGGTTTTCCTTATTCGTCGGAACTCTTTTGATGAGC 226

Qy     209 CAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAAAGTCCAGCACTC 268
           |||  |  | ||  |    |   ||   |    |||          ||  |  |||  ||
Db     227 AAAATGAAATAAGTGGTCGGTATTCTTGGAACTTCCAGAAATCCACCAGTTCCTTCAACT 286

Qy     269 TGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCCTCGTGGGTGGTT 328
           | | | |  |           |    ||| | || || |||| |||     |||  |
Db     287 TCACCATCACAGCCTCACAAGTCGTGGACTCAGCAGTATACTTCTGTGCTCTGGACAGCA 346

Qy     329 CTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGA 388
           |||     |  | | |||||||||| |||||   | | |   |   ||  |||||||||
Db     347 GTGCTTCCAAGATAATCTTTGGATCAGGGACCAGACTCAGCATCCGGCCAAATATCCAGA 406

Qy     389 ACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCC 448
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407 ACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCC 466

Qy     449 TATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATA 508
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     467 TATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATA 526

Qy     509 TCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGG 568
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 TCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGG 586

Qy     569 CCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAG 628
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     587 CCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAG 646

Qy     629 AAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCT 688
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     647 AAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCT 706

Qy     689 TTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCC 748
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     707 TTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCC 766

Qy     749 TCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGA 804
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     767 TCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGA 822
```

```
RESULT   2
5189147-2
;Patent No. 5189147
;    APPLICANT: SAITO, HARUO;KRANZ, DAVID M.;ELSEN, HERMAN N.;
;TONEGAWA, SUSUMU
;    TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR
;ANTIBODY
;    NUMBER OF SEQUENCES: 21
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/271,216
;      FILING DATE: 14-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 666,988
;      FILING DATE: 31-OCT-1984
;      APPLICATION NUMBER: 620,122
;      FILING DATE: 13-JUN-1984
;SEQ ID NO:2:
;      LENGTH: 1372
5189147-2
```

Query Match 22.3%; Score 299.6; DB 6; Length 1372;
Best Local Similarity 69.2%; Pred. No. 5.1e-79;
Matches 429; Conservative 0; Mismatches 179; Indels 12; Gaps 1;

```
Qy     259 TCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCCTC 318
           |||  |   ||| |     ||   ||   |  |||||||  | ||||||||| |||
Db     329 tccttccacctgcggaaagcctccgtgcactggagcgactcggctgtgtacttctgtgct 388

Qy     319 GTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCT 378
           ||| | || | ||||||    ||||| ||||||||||| ||| || | | |||| ||||
Db     389 gtgagcggctttgcaagtgcgctgacatttggatctggcacaaaagtcattgttctacca 448

Qy     379 GATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAG 438
           | |||||||||||| || ||||| |||||||| | | |||  ||   ||    ||||
Db     449 tacatccagaacccagaacctgctgtgtaccagttaaaagatcctcggtctcaggacagc 508

Qy     439 TCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCT 498
           |  ||||||| |||||||| ||||| || ||||  ||||||  |  |||  | ||| |||
Db     509 accctctgcctgttcaccgactttgactcccaaatcaatgtgccgaaaaccatggaatct 568

Qy     499 GATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAAC 558
           |    ||  ||||| |||||||||||||| |||||||   ||||||| | |||||||||
Db     569 ggaacgttcatcactgacaaaactgtgctggacatgaaagctatggattccaagagcaat 628

Qy     559 AGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGC 618
           | ||  | ||||||||||||| |  |   |||  | ||   | |   ||||||  | | |
Db     629 ggggccattgcctggagcaaccagacaagcttcacctgccaagatatcttcaaagagacc 688

Qy     619 ATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTC 678
           |            || ||  || ||||||  ||||   | |||||||| || | ||  |
Db     689 a------------acgccacctaccccagttcagacgttccctgtgatgccacgttgacc 736

Qy     679 GAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC 738
           |||||||||||||||||||| ||||||||||||||||||||||||||| || ||  ||
Db     737 gagaaaagctttgaaacagatatgaacctaaactttcaaaacctgtcagttatgggactc 796

Qy     739 CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCC 798
           |||||||||| |||||||| || || ||||| |||||||||||||||| ||||||||||
Db     797 cgaatcctcctgctgaaagtagcgggatttaacctgctcatgacgctgaggctgtggtcc 856

Qy     799 AGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGC 858
           || |||| |||||||||| ||  |||        | |    |  |  | |||||  |  | |
Db     857 agttgaggtctgcaagactgacagagcctgactcccaagttccgtcctcctcacccctcc 916

Qy     859 CCCTCTTCTCCCTCTCCAAA 878
           |  | ||| |    | |||
Db     917 gctccctcttcaagccaaaa 936
```

```
RESULT   3
US-09-082-593-9
; Sequence 9, Application US/09082593
```

```
; Patent No. 6180104
; GENERAL INFORMATION:
;  APPLICANT: DAVIS, MARK M.
;  APPLICANT: HEDRICK, STEPHEN M.
;  TITLE OF INVENTION: T CELL RECEPTOR BETA SUBUNIT
;  FILE REFERENCE: JX1193-195DIV2
;  CURRENT APPLICATION NUMBER: US/09/082,593
;  CURRENT FILING DATE: 1998-05-20
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12)..(821)
;   FEATURE:
;   NAME/KEY: Signal peptide
;   LOCATION: (12)..(71)
;   FEATURE:
;   NAME/KEY: Mature peptide
;   LOCATION: (72)..(821)
US-09-082-593-9
```

Query Match 19.7%; Score 264.4; DB 4; Length 990;
Best Local Similarity 68.4%; Pred. No. 1.1e-68;
Matches 386; Conservative 0; Mismatches 166; Indels 12; Gaps 1;

```
Qy    315 CCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTT 374
Db    347 cctttatgggggcagtggcaacaagctcatctttggaactggcactctgctttctgtcaa 406

Qy    375 ACCTGATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGA 434
Db    407 gccaaacatccagaacccagaacctgctgtgtaccagttaaaagatcctcggtctcagga 466

Qy    435 CAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGA 494
Db    467 cagcaccctctgcctgttcaccgactttgactcccaaatcaatgtgccgaaaaccatgga 526

Qy    495 TTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAG 554
Db    527 atctggaacgttcatcactgacaaaactgtgctggacatgaaagctatggattccaagag 586

Qy    555 CAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAA 614
Db    587 caatggggccattgcctggagcaaccagacaagcttcacctgccaagatatcttcaaaga 646

Qy    615 CAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCT 674
Db    647 gacca-----acgccacctaccccagttcagacgttccctgtgatgccacgtt 694

Qy    675 GGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGG 734
Db    695 gaccgagaaaagctttgaaacagatatgaacctaaactttcaaaacctgtcagttatggg 754

Qy    735 GTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTG 794
Db    755 actccgaatcctcctgctgaaagtagcgggatttaacctgctcatgacgctgaggctgtg 814

Qy    795 GTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCA 854
Db    815 gtccagttgaggtctgcaagactgacagagcctgactcccaagttccgtcctcctcaccc 874

Qy    855 TTGCCCCTCTTCTCCCTCTCCAAA 878
Db    875 ctccgctccctcttcaagccaaaa 898
```

RESULT 4
US-08-416-336-1

```
; Sequence 1, Application US/08416336
; Patent No. 5807714
;  GENERAL INFORMATION:
;    APPLICANT:  Ishizaka, Kimishige
;    APPLICANT:  Ishii, Yasuyuki
;    TITLE OF INVENTION:  METHOD OF PRODUCTION OF ANTIGEN-SPECIFIC
;    TITLE OF INVENTION:  GLYCOSYLATION INHIBITING FACTOR
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/416,336
;      FILING DATE:  04-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa H.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07246/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
US-08-416-336-1
```

Query Match 18.4%; Score 247.4; DB 1; Length 804;
Best Local Similarity 72.6%; Pred. No. 1.1e-63;
Matches 339; Conservative 0; Mismatches 116; Indels 12; Gaps 1;

```
Qy    338 AACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCTGATATCCAGAACCCTGACC 397
Db    350 AACTGACATTTGGGAAAGGAACTCTCTTAACTGTGACTCCAAACATCCAGAACCCAGAAC 409

Qy    398 CTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCG 457
Db    410 CTGCTGTGTACCAGTTAAAAGATCCTCGGTCTCAGGACAGCACCCTCTGCCTGTTCACCG 469

Qy    458 ATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACA 517
Db    470 ACTTTGACTCCCAAATCAATGTGCCGAAAACCATGGAATCTGGAACGTTCATCACTGACA 529

Qy    518 AAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCA 577
Db    530 AAACTGTGCTGGACATGAAAGCTATGGATTCCAAGAGCAATGGGGCCATTGCCTGGAGCA 589

Qy    578 ACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCT 637
Db    590 ACCAGACAAGCTTCACCTGCCAAGATATCTTCAAAGAGACCA-----ACGCCA 637

Qy    638 TCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAG 697
Db    638 CCTACCCCAGTTCAGACGTTCCCTGTGATGCCACGTTGACCGAGAAAAGCTTTGAAACAG 697

Qy    698 ATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAG 757
Db    698 ATATGAACCTAAACTTTCAAAACCTGTCAGTTATGGGACTCCGAATCCTCCTGCTGAAAG 757
```

```
Qy    758 TGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGA 804
          | || || ||||| |||||||||||||||| |||||||||||| |||
Db    758 TAGCGGGATTTAACCTGCTCATGACGCTGAGGCTGTGGTCCAGTTGA 804


RESULT   5
US-09-140-084-25
; Sequence 25, Application US/09140084A
; Patent No. 6300065
; GENERAL INFORMATION:
;  APPLICANT: Kieke, et al.
;  TITLE OF INVENTION: Yeast Cell Surface Display of Proteins and Uses Thereof
;  FILE REFERENCE: D6061CIP2
;  CURRENT APPLICATION NUMBER: US/09/140,084A
;  CURRENT FILING DATE: 1998-08-26
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Sequence
;   OTHER INFORMATION: showing mutations in T-cell receptor
US-09-140-084-25


  Query Match             4.3%;  Score 57;  DB 4;  Length 747;
  Best Local Similarity   65.1%;  Pred. No. 2.7e-07;
  Matches   84;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    259 TCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCCTC 318
          |||  |   ||| |     ||   ||   |  ||||||| | ||||||||||| |||
Db    619 tccttccacctgcggaaagcctccgtgcactggagcgactcggctgtgtacttctgtgct 678

Qy    319 GTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCT 378
          ||| | || | ||||||    ||||| |||||||||||| ||| || | | |||| ||||
Db    679 gtgagcggctttgcaagtgcgctgacatttggatctggcacaaaagtcattgttctacca 738

Qy    379 GATATCCAG 387
           | ||| ||
Db    739 tacatctag 747


RESULT   6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;     ·CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.2%;  Score 55.8;  DB 1;  Length 7218;
  Best Local Similarity   7.5%;  Pred. No. 2e-06;
  Matches   33;  Conservative  222;  Mismatches  184;  Indels    0;  Gaps    0;

Qy    712 TTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGTGGCCGGGTTTAAT 771
          |||   |  |  |||  || |   ||||||      || | |   | ||  |  |   ||
Db   1007 TTTGCCATACGCTCACAGAATTAATTCCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGAT 1066

Qy    772 CTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGT 831
          :: ::: :  : :: :  :: :  :::  ::    ::: :    :: :    :  ::: :
Db   1067 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1126

Qy    832 GCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCT 891
           ::::::: :::::::::::: : :: ::::::::::::::::::::    :        ::::
Db   1127 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1186

Qy    892 CCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAA 951
          ::: :::::      :    :: :: :: ::::: : :::::::  :  : :: ::
Db   1187 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1246

Qy    952 CTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCAC 1011
          ::   :::: :::   ::: :  ::    :: ::          :: ::    : :: :: :: :
Db   1247 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1306

Qy   1012 CCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTG 1071
          ::::::: :::::::: :: :: ::: :: :: :::  : ::: :  :      :    ::
Db   1307 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1366

Qy   1072 CTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATC 1131
          :: :  ::::::::  :: : : : ::::::::: ::::::     :: ::::: :: ::
Db   1367 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1426

Qy   1132 CCACAGATGATGGATCTTC 1150
          :: :   :|    |  |||
Db   1427 YYYYYYYYGTACCAAATTC 1445


RESULT   7
US-08-411-098-6
; Sequence 6, Application US/08411098
; Patent No. 5830755
;  GENERAL INFORMATION:
;    APPLICANT:  HWU, PATRICK; NISHIMURA,
;    APPLICANT:  MICHAEL; ROSENBERG, STEVEN A.
;    TITLE OF INVENTION:  T-CELL RECEPTORS AND
;    TITLE OF INVENTION:  THEIR USE IN THERAPEUTIC AND DIAGNOSTIC
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
```

; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,098
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4188
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: UNKNOWN
US-08-411-098-6

Query Match 3.1%; Score 41.8; DB 2; Length 93;
Best Local Similarity 65.6%; Pred. No. 0.0029;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 304 GTGTACTACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAA 363
Db 1 GAGTACTTCTGTGCTGTGGGTGCCACCGGTAACCAGTTCTATTTTGGGACAGGGACAAGT 60

Qy 364 TTGACTGTTTTACCTGATATCCAGAACCCTGAC 396
Db 61 TTGACGGTCATTCCAAATATCCAGAACCCTGAC 93

RESULT 8
US-08-480-753-7
; Sequence 7, Application US/08480753
; Patent No. 5830675
; GENERAL INFORMATION:
; APPLICANT: Targan M.D., Stephan R.
; APPLICANT: Vidrich Ph.D., Alda M.
; TITLE OF INVENTION: METHODS FOR SELECTIVELY DETECTING
; TITLE OF INVENTION: PERINUCLEAR ANTI-NEUTROPHIL CYTOPLASMIC ANTIBODY OF
; TITLE OF INVENTION: ULCERATIVE COLITIS, PRIMARY SCLEROSING CHOLANGITIS, OR
; TITLE OF INVENTION: TYPE I AUTOIMMUNE HEPATITIS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wendy A. Whiteford, Esq.
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,753
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whiteford, Wendy A.
; REGISTRATION NUMBER: 36,964
; REFERENCE/DOCKET NUMBER: P07 33571
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Gut-associated lymphoid
; CELL TYPE: Lymphocyte
; IMMEDIATE SOURCE:
; CLONE: 5-4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..645
; OTHER INFORMATION: /codon_start= 1
; OTHER INFORMATION: /product= "Kappa Light Chain of ANCA associated
; OTHER INFORMATION: with Ulcerative Colitis"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 1..3
; OTHER INFORMATION: /label= N-TerminalTag
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 4..285
; OTHER INFORMATION: /label= VKSEGMENT
; OTHER INFORMATION: /note= ""VKSEGMENT" refers to Variable Segment of
; OTHER INFORMATION: the Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 286..327
; OTHER INFORMATION: /label= JK
; OTHER INFORMATION: /note= ""JK" refers to Joining Segment of the
; OTHER INFORMATION: Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 328..645
; OTHER INFORMATION: /label= CK
; OTHER INFORMATION: /note= ""CK" refers to Constant Segment of the
; OTHER INFORMATION: Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 4..66
; OTHER INFORMATION: /label= FR1
; OTHER INFORMATION: /note= ""FR1" refers to Framework Region 1"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 67..102
; OTHER INFORMATION: /label= CDR1
; OTHER INFORMATION: /note= ""CDR1" refers to Complimentarity
; OTHER INFORMATION: Determining Region 1"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 103..147
; OTHER INFORMATION: /label= FR2
; OTHER INFORMATION: /note= ""FR2" refers to Framework Region 2"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 148..168
; OTHER INFORMATION: /label= CDR2
; OTHER INFORMATION: /note= ""CDR2" refers to Complimentarity
; OTHER INFORMATION: Determining Region 2"
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: 169..264
; OTHER INFORMATION: /label= FR3

```
;      OTHER INFORMATION:  /note= ""FR3" refers to Framework Region 3"
;    FEATURE:
;      NAME/KEY:  misc_RNA
;      LOCATION:  265..294
;      OTHER INFORMATION:  /label= CDR3
;      OTHER INFORMATION:  /note= ""CDR3" refers to Complimentarity
;      OTHER INFORMATION:  Determining Region 3"
;    FEATURE:
;      NAME/KEY:  misc_RNA
;      LOCATION:  295..327
;      OTHER INFORMATION:  /label= FR4
;      OTHER INFORMATION:  /note= ""FR4" refers to Framework Region 4"
US-08-480-753-7


  Query Match             3.0%;  Score 39.8;  DB 2;  Length 645;
  Best Local Similarity   45.6%;  Pred. No. 0.031;
  Matches  140;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy      70 CCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAAC 129
           ||   | || ||   |  | | |||  ||  ||  | | |  |||| ||   ||   |
Db      19 CCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAG 78

Qy     130 ATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTT 189
             || ||  |  |  |    |  | ||||| || |||    |  |||| |  ||  |  |
Db      79 GGTGTTAGCAGCGGCTCCTTAGCCTGGTATCAGCAGAAAGCTGGCCAGGCTCCCAGGCTC 138

Qy     190 ATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
           | ||  | ||       |  |  |  ||   |      | |    |   |  |  ||
Db     139 CTCATCTATGGTGCATCCAGGAGGGCCACTGGCATCCCAGACAGGTTCACTGGCAGTGGG 198

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
             | |    |  |||||| | | |  || |  |    |     ||   ||| |||||
Db     199 TCTGGGACAGACTTCACTCTCACCATCACCAGACTGGAGCCTGAAGATTTTGCAGTGTAT 258

Qy     310 TACTGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACT 369
           |||||  |    |  ||||    |   |  | | || || ||  |||||||  || ||  |
Db     259 TACTGTCAGCAGTATGGTAGCTCCCAGGGATTCACTTTCGGCCCTGGGACCAAAGTGGAT 318

Qy     370 GTTTTAC 376
           |   ||
Db     319 CTCAAAC 325


RESULT   9
US-08-591-632-43
; Sequence 43, Application US/08591632
; Patent No. 6261558
;  GENERAL INFORMATION:
;    APPLICANT:  Barbas, Carlos F.
;    APPLICANT:  Burton, Dennis R.
;    APPLICANT:  Lerner, Righard A.
;    TITLE OF INVENTION:  SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL
;    TITLE OF INVENTION:  ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10550 No. 6261558th Torrey Pines Road, TPC 8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,632
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/11907
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,841
;      FILING DATE:  19-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,619
;      FILING DATE:  26-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/139,409
;      FILING DATE:  19-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 332.3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 784-2937
;      TELEFAX:  (619) 784-9399
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4691 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
US-08-591-632-43


  Query Match             2.8%;  Score 37.2;  DB 4;  Length 4691;
  Best Local Similarity   46.5%;  Pred. No. 0.52;
  Matches  120;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy      70 CCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAAC 129
           ||   | || ||   |  | | |||  ||  ||  | | |  |||| ||   ||   ||
Db    2692 CCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAC 2751

Qy     130 ATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTT 189
           | || ||  |  |  ||  |  | |||||||| |||   ||  |||| |  ||  |  |
Db    2752 AGTGTTAGCAGGGCCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTC 2811

Qy     190 ATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
           | ||  | ||       |  |  |  ||   |      | |    |   |  |  ||
Db    2812 CTCATCTATGGTACATCCAGCAGGGCCACTGGCATCCCAGACAGGTTCAGTGGCAGTGGG 2871

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
             | |    |  |||||| | | |  | |  |    |     ||   ||| ||||||
Db    2872 TCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTAC 2931

Qy     310 TACTGCCTCGTGGGTGGT 327
           |||||  |    |  ||||
Db    2932 TACTGTCAGCAGTATGGT 2949


RESULT  10
US-08-591-632-51
; Sequence 51, Application US/08591632
; Patent No. 6261558
;  GENERAL INFORMATION:
;    APPLICANT:  Barbas, Carlos F.
;    APPLICANT:  Burton, Dennis R.
;    APPLICANT:  Lerner, Righard A.
;    TITLE OF INVENTION:  SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL
;    TITLE OF INVENTION:  ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10550 No. 6261558th Torrey Pines Road, TPC 8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/591,632
;       FILING DATE:  19-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/11907
;       FILING DATE:  19-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/308,841
;       FILING DATE:  19-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/233,619
;       FILING DATE:  26-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/139,409
;       FILING DATE:  19-OCT-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fitting, Thomas
;       REGISTRATION NUMBER:  34,163
;       REFERENCE/DOCKET NUMBER:  TSRI 332.3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 784-2937
;       TELEFAX:  (619) 784-9399
;   INFORMATION FOR SEQ ID NO:  51:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6166 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
US-08-591-632-51


  Query Match             2.8%;  Score 37.2;  DB 4;  Length 6166;
  Best Local Similarity   46.5%;  Pred. No. 0.59;
  Matches  120;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy      70 CCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAAC 129
           ||   | || ||    |  | |  ||| ||  ||  | | |  |||| ||   ||   ||
Db    4622 CCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAC 4681

Qy     130 ATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTT 189
           | || ||  |  |  ||  |  | |||||||| |||   ||  |||| |  ||  |  |
Db    4682 AGTGTTAGCAGGGCCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTC 4741

Qy     190 ATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
             | ||  | ||       |  |  |  ||   |       | |     |   |  | ||
Db    4742 CTCATCTATGGTACATCCAGCAGGGCCACTGGCATCCCAGACAGGTTCAGTGGCAGTGGG 4801

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
           | |    |  |||||| | | |   | |  |     |     ||   ||| ||||||
Db    4802 TCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTAC 4861

Qy     310 TACTGCCTCGTGGGTGGT 327
           |||||  |   |  ||||
Db    4862 TACTGTCAGCAGTATGGT 4879


RESULT  11
US-09-025-769B-46
; Sequence 46, Application US/09025769B
; Patent No. 6300064
;  GENERAL INFORMATION:
;     APPLICANT:  Knappik, Achim
;     APPLICANT:  Pack, Peter
;     APPLICANT:  Ilag, Vic
;     APPLICANT:  Ge, Liming
;     APPLICANT:  Moroney, Simon
;     APPLICANT:  Plueckthun, Andreas
;     TITLE OF INVENTION:  Protein/(Poly)peptide libraries
;     NUMBER OF SEQUENCES:  373
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  James F. Haley, Jr., Esq. c/o Fish & Neave
;       STREET:  1251 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10021
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/025,769B
;       FILING DATE:  18-FEB-1998
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 95 11 3021.0
;       FILING DATE:  18-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  James F. Haley, Jr., Esq.
;       REGISTRATION NUMBER:  27,794
;       REFERENCE/DOCKET NUMBER:  MORPHO/5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212)596-9000
;       TELEFAX:  (212)596-9090
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  330 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "synthetic gene"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..330
;       OTHER INFORMATION:  /product= "V kappa 3"
US-09-025-769B-46


  Query Match             2.7%;  Score 36.8;  DB 4;  Length 330;
  Best Local Similarity   47.1%;  Pred. No. 0.17;
  Matches  113;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy      77 CCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTA 136
           || ||   |   | |   ||  ||   | || | | | ||| ||    | | | |  | |
Db      29 CCCTGAGCCTGTCTCCGGGCGAACGTGCGACCCTGAGCTGCAGAGCGAGCCAGAGCGTGA 88

Qy     137 CAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTC 196
             |     ||| |  |||||||||| |||   ||  | |||| ||| ||  |  | |||
Db      89 GCAGCAGCTATCTGGCGTGGTACCAGCAGAAACCAGGTCAAGCACCGCGTCTATTAATTT 148

Qy     197 AAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAA 256
           | ||    |  |     |  ||    |    |||     |       | |||   | | |
Db     149 ATGGCGCGAGCAGCCGTGCAACTGGGGTCCCGGCGCGTTTTAGCGGCTCTGGATCCGGCA 208

Qy     257 AGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCC 316
           |      || |||| | |   | |  |     |     |||  ||| ||||| || ||||
Db     209 CGGATTTTACCCTGACCATTAGCAGCCTGGAACCTGAAGACTTTGCGGTGTATTATTGCC 268


RESULT  12
US-08-619-554-1
; Sequence 1, Application US/08619554
; Patent No. 5821353
;  GENERAL INFORMATION:
;     APPLICANT:  DOUGLAS, Cameron M.
;     APPLICANT:  CHREBET, Gary L.
;     APPLICANT:  CLEMAS, Joseph
;     APPLICANT:  EL-SHERBEINI, Mohammed
;     APPLICANT:  FOOR, Forrest
;     APPLICANT:  KAHN, Jennifer,
;     APPLICANT:  KELLY, Rosemarie, - PARENT, S.A.
```

```
;    APPLICANT:  MARRINAN, Jean, - RAMADAN, N.M.
;    APPLICANT:  MORIN, Nancy, - REGISTER, E.A
;    APPLICANT:  ONISHI, Janet, - SHEI, Gan-Ju
;    TITLE OF INVENTION:  DNA ENCODING 1,3 BETA-D GLUCAN
;    TITLE OF INVENTION:  SYNTHASE SUBUNITS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JOSEPH A. COPPOLA - MERCK & CO., INC.
;      STREET:  126 EAST LINCOLN AVENUE - P.O. BOX 2000
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,554
;      FILING DATE:  01-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COPPOLA, JOSEPH A
;      REGISTRATION NUMBER:  38,413
;      REFERENCE/DOCKET NUMBER:  19104PI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732-594-6734
;      TELEFAX:  732-594-4720
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-619-554-1


  Query Match             2.7%;  Score 36.4;  DB 1;  Length 7655;
  Best Local Similarity   74.2%;  Pred. No. 1.1;
  Matches   46;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    1125 CGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAAT 1184
           | ||||   | | ||| | |||| |     |||||||||| |||||||||| ||||||
Db    1700 CTCCATGAGAAATATGTTTGATCATTTTATGGTTCTCTTGGACTCTAGGTCCTCGAGAAT 1759

Qy    1185 GT 1186
           ||
Db    1760 GT 1761


RESULT  13
US-08-405-034-3
; Sequence 3, Application US/08405034
; Patent No. 5744585
; Patent No. 5744585 5712371
;  GENERAL INFORMATION:
;    APPLICANT:  Medenica, Rajko D.
;    APPLICANT:  Mukerjee, Sonjoy
;    TITLE OF INVENTION:  Human Monoclonal Antibody Against Lung
;    TITLE OF INVENTION:  Carcinoma
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DeWitt Ross & Stevens, S.C.
;      STREET:  8000 Excelsior Drive, Suite 401
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53717-1914
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,034
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34656.009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-831-2100
;      TELEFAX:  608-831-2106
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FRAGMENT TYPE:  C-terminal
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..402
US-08-405-034-3


  Query Match             2.7%;  Score 35.6;  DB 1;  Length 402;
  Best Local Similarity   45.2%;  Pred. No. 0.42;
  Matches  131;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy      77 CCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAACATTGCTA 136
           || ||   |  | |  |||  ||  ||  | | |  |||| ||   ||  | | |  ||
Db      89 CCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTTTTA 148

Qy     137 CAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTTATTATTC 196
             |     |   |  | |||||||| |||   ||  |||| |  ||  |  |  | ||
Db     149 GCAGAAGCTTCTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGCCTCCTCATCT 208

Qy     197 AAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAA 256
           | ||       || |  | |||   |       | |    |   |  |  ||      | |
Db     209 ATGGTGCATCCACCAGGGCTACTGGCATCCCAGACAGGTTCAGTGGCAGTGGGTCTGGGA 268

Qy     257 AGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTACTACTGCC 316
             | || ||| | | |   | |  |     |      ||   ||| ||||| ||||| |
Db     269 CAGACTTCATTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTC 328

Qy     317 TCGTGGGTGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTG 366
             |  ||||    ||   |    || |||||    |||||  |  ||
Db     329 AGCAGTATGGTAGCTCAGCTCGGTACACTTTTGGCCAGGGGACCAAGCTG 378


RESULT  14
US-08-300-386A-2
; Sequence 2, Application US/08300386A
; Patent No. 5667988
;  GENERAL INFORMATION:
;    APPLICANT:  Barbas, Carlos F,III
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  METHODS FOR PRODUCING ANTIBODY LIBRARIES
;    TITLE OF INVENTION:  USING UNIVERSAL OR RANDOMIZED IMMUNOGLOBULIN LIGHT CHAINS
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute
;      STREET:  10666 No. 5667988th Torrey Pines Road, TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
```

```
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,386A
;      FILING DATE:  02-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/174,674
;      FILING DATE:  28-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/826,623
;      FILING DATE:  27-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/012,566
;      FILING DATE:  02-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 409.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  646 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-300-386A-2


  Query Match             2.7%;  Score 35.6;  DB 1;  Length 646;
  Best Local Similarity   46.1%;  Pred. No. 0.54;
  Matches  119;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy      70 CCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAAC 129
           ||   | || ||   |  | |  |||  ||  ||  | | |  |||| ||   ||   ||
Db      16 CCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAC 75

Qy     130 ATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTT 189
           | || ||  |  |  ||  |  | ||||||||| |||   ||  |||| |  ||  |  |
Db      76 AGTGTTAGCAGGGCCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTC 135

Qy     190 ATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
           | ||  | ||       |  |  |  ||   |      ||        |  |  ||
Db     136 CTCATCTATGGTACATCCAGCAGGGCCACTGGCATCCCAGACAGGTCCAGTGGCAGTGGG 195

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
           | |    |  |||||| | | |   | |  |    |     ||   ||| ||||||
Db     196 TCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTAC 255

Qy     310 TACTGCCTCGTGGGTGGT 327
           |||||  |    |  ||||
Db     256 TACTGTCAGCAGTATGGT 273


RESULT  15
US-08-931-645-2
; Sequence 2, Application US/08931645
; Patent No. 6096551
;  GENERAL INFORMATION:
;    APPLICANT:  Barbas, Carlos F,III
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  METHODS FOR PRODUCING ANTIBODY LIBRARIES
;    TITLE OF INVENTION:  USING UNIVERSAL OR RANDOMIZED IMMUNOGLOBULIN LIGHT CHAINS
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute
;      STREET:  10666 No. 6096551th Torrey Pines Road, TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,645
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,386
;      FILING DATE:  02-SEP-1994
;      APPLICATION NUMBER:  US 08/174,674
;      FILING DATE:  28-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/826,623
;      FILING DATE:  27-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/012,566
;      FILING DATE:  02-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 409.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  646 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-931-645-2


  Query Match             2.7%;  Score 35.6;  DB 3;  Length 646;
  Best Local Similarity   46.1%;  Pred. No. 0.54;
  Matches  119;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy      70 CCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATAACCTGTAGCCACAACAAC 129
           ||   | || ||   |  | |  |||  ||  ||  | | |  |||| ||   ||   ||
Db      16 CCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAC 75

Qy     130 ATTGCTACAAATGATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGACCACGATTT 189
           | || ||  |  |  ||  |  | ||||||||| |||   ||  |||| |  ||  |  |
Db      76 AGTGTTAGCAGGGCCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTC 135

Qy     190 ATTATTCAAGGATACAAGACAAAAGTTACAAACGAAGTGGCCTCCCTGTTTATCCCTGCC 249
           | ||  | ||       |  |  |  ||   |      ||        |  |  ||
Db     136 CTCATCTATGGTACATCCAGCAGGGCCACTGGCATCCCAGACAGGTCCAGTGGCAGTGGG 195

Qy     250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGTGTAC 309
           | |    |  |||||| | | |   | |  |    |     ||   ||| ||||||
Db     196 TCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTAC 255
```

```
Qy     310 TACTGCCTCGTGGGTGGT 327
           ||||| |    |  ||||
Db     256 TACTGTCAGCAGTATGGT 273
```

Search completed: June 6, 2002, 07:38:13
Job time: 5315 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:39:08 ; Search time 2149.71 Seconds
(without alignments)
379.649 Million cell updates/sec

Title: US-08-881-509-5
Perfect score: 39
Sequence: 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	100.0	39	6	A93131	A93131 Sequence 5
2	39	100.0	1318	9	HSTCRJUNC	X98410 H.sapiens m
3	31	79.5	42	9	HSA235208	AJ235208 Homo sapi
4	30.4	77.9	48	9	HSU27254	U27254 Human isola
5	29.8	76.4	222	9	AF333590	AF333590 Homo sapi
6	29	74.4	51	9	AF043888	AF043888 Homo sapi
7	29	74.4	183	9	HUMTCAZA	M17666 Human T-cel
8	28.4	72.8	51	9	AF043877	AF043877 Homo sapi
9	28.4	72.8	51	9	AF043886	AF043886 Homo sapi
10	28.4	72.8	51	9	AF043890	AF043890 Homo sapi
11	28.2	72.3	39	9	HSU30448	U30448 Human isola
12	28.2	72.3	63	9	HSU30390	U30390 Human isola
13	28.2	72.3	645	9	HSTCRAJ3	X02886 Human gene
14	28.2	72.3	71153	9	HUAE000662	AE000662 Homo sapi
15	28.2	72.3	97630	9	HUMTCRADCV	M94081 Human Tcr-C
16	28.2	72.3	175053	2	AC023226	AC023226 Homo sapi
17	28	71.8	39	9	HSU30428	U30428 Human isola
18	28	71.8	51	9	AF043875	AF043875 Homo sapi
19	28	71.8	51	9	AF043884	AF043884 Homo sapi
20	28	71.8	411	9	HUMTCVJCC	M97706 Human T-cel
21	28	71.8	716	9	S63879	S63879 TCR V alpha
22	27.8	71.3	432	9	HUMIGTCACA	M81774 Homo sapien
23	27.6	70.8	36	6	A93133	A93133 Sequence 7
24	27	69.2	51	9	AF043874	AF043874 Homo sapi
25	27	69.2	51	9	AF043876	AF043876 Homo sapi
26	27	69.2	51	9	AF043892	AF043892 Homo sapi
27	27	69.2	51	9	AF043894	AF043894 Homo sapi
28	27	69.2	87	9	HUMTCRACG	L42801 Homo sapien
29	26.6	68.2	98	9	S69140	S69140 Homo sapien
30	26.2	67.2	1341	6	A93127	A93127 Sequence 1
31	25	64.1	214	9	HSU40776	U40776 Human T-cel
32	24.8	63.6	42	9	HSTCELL14	X81549 H.sapiens r
c 33	24.6	63.1	46255	3	U20861	U20861 Caenorhabdi
34	24	61.5	51	9	HSU14029	U14029 Human T cel
35	24	61.5	234	9	AF041171	AF041171 Homo sapi
36	24	61.5	246	9	HSTCARAVB	Z22966 H.sapiens m
c 37	23.2	59.5	136960	9	AC092595	AC092595 Homo sapi
38	23.2	59.5	150899	2	AC018400	AC018400 Homo sapi
39	23.2	59.5	166598	2	AC106723	AC106723 Homo sapi
c 40	23.2	59.5	167026	9	AP002783	AP002783 Homo sapi
41	23.2	59.5	170603	2	AC069289	AC069289 Homo sapi
42	23.2	59.5	175810	2	AC068727	AC068727 Homo sapi
43	23	59.0	134640	9	AC027130	AC027130 Homo sapi
44	23	59.0	218939	2	AC087558	AC087558 Mus muscu
c 45	22.8	58.5	937	9	HSA326131	AJ326131 Homo sapi

ALIGNMENTS

RESULT 1
A93131
LOCUS A93131 39 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent EP0816496.
ACCESSION A93131
VERSION A93131.1 GI:6741520
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 39)
AUTHORS Schendel,D.J.
TITLE T-cells specific for kidney carcinoma
JOURNAL Patent: EP 0816496-A 5 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES Location/Qualifiers
source 1. .39
/organism="unidentified"
/db_xref="taxon:32644"
CDS <1. .>39
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69531.1"

```
                     /db_xref="GI:6741521"
                     /translation="CLVLSGSARQLTF"
BASE COUNT         5 a      12 c       8 g      14 t
ORIGIN


  Query Match             100.0%;  Score 39;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-06;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
          |||||||||||||||||||||||||||||||||||||||
Db      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39


RESULT   2
HSTCRJUNC
LOCUS       HSTCRJUNC                1318 bp    mRNA    linear   PRI 08-JAN-1997
DEFINITION  H.sapiens mRNA for rearranged TCR junctional sequences.
ACCESSION   X98410
VERSION     X98410.1  GI:1770560
KEYWORDS    J gene; junction; T cell receptor; TCR junctional sequence; V gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1318)
  AUTHORS   Jantzer,P.U. and Schendel,D.J.
  TITLE     Tumor-infiltrating lymphocytes recognizing spontaneously arising
            renal cell carcinomas express T cell receptors characteristic of a
            secondary immune response
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1318)
  AUTHORS   Jantzer,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology,
            Goethestr. 31, D- 80336 Munich, FRG
FEATURES             Location/Qualifiers
     source          1. .1318
                     /organism="Homo sapiens"
                     /isolate="patients 22 and 26"
                     /db_xref="taxon:9606"
                     /rearranged
     CDS             <1. .>1318
                     /note="V gene/J gene junction"
                     /codon_start=1
                     /product="TCR junctional sequence"
                     /protein_id="CAA67057.1"
                     /db_xref="GI:1770561"
                     /translation="CLVGGSARQLTFGXCLATGSARQLTFGXCLVLSGSARQLTFGXC
                     LVLSGSARQLTFGXCLVLSGSARQLTFGXCLVGGSARQLTFGXCLVGGSARQLTFGXC
                     LVSGSARQLTFGXCLDSGSARQLTFGXCLDSGSARQLTFGXCLPSGSARQLTFGXCLA
                     TGSARQLTFGXCLVGGDTDKLIFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLV
                     VSGSARQLTFGXCLALGGSARQLTFGXCLALAGSARQLTFGXCLAPSGSARQLTFGXC
                     LVAGGGNTPLVFGXCLVGGSARQLTFGXCLVGGSARQLTFGXCLVGRSARQLTFGXCL
                     ATGSARQLTFGXCLVGAGGYQKVTFGXCLVLSGSARQLTFGXCLVAPGGATNKLIFGX
                     CLVVLGGSQGNLIFGXCLVGPNNAGNMLTFGXCLVATNQGRNCSDLW"
BASE COUNT       206 a    329 c    352 g    344 t     87 others
ORIGIN


  Query Match             100.0%;  Score 39;  DB 9;  Length 1318;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
          |||||||||||||||||||||||||||||||||||||||
Db     85 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 123


RESULT   3
HSA235208
LOCUS       HSA235208                  42 bp    mRNA    linear   PRI 12-MAR-1999
DEFINITION  Homo sapiens mRNA for T cell receptor alpha chain V-J junctional
            region (TCRAV7AJ16S3).
ACCESSION   AJ235208
VERSION     AJ235208.1  GI:3851223
KEYWORDS    T cell receptor; T cell receptor alpha chain; variable region.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Cross,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1998) Cross S.M., Queensland Institute of Medical
            Research, The Bancroft Centre, 300 Herston Road, Brisbane,
            AUSTRALIA 4029
REFERENCE   2  (bases 1 to 42)
  AUTHORS   Misko,I.S., Cross,S.M., Khanna,R., Elliott,S.L., Schmidt,C.,
            Pye,S.J. and Silins,S.L.
  TITLE     Crossreactive recognition of viral, self, and bacterial peptide
            ligands by human class I-restricted cytotoxic T lymphocyte
            clonotypes: implications for molecular mimicry in autoimmune
            disease
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (5), 2279-2284 (1999)
  MEDLINE   99162595
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /haplotype="A1, A2, B8, B45, Bw6"
                     /rearranged
                     /tissue_type="blood"
                     /cell_type="cytotoxic T lymphocyte"
                     /cell_line="SP1"
     V_region        1. .42
BASE COUNT         6 a      9 c     11 g     16 t
ORIGIN


  Query Match             79.5%;  Score 31;  DB 9;  Length 42;
  Best Local Similarity   87.2%;  Pred. No. 0.0063;
  Matches   34;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
          |||   ||  ||||||||||||||||||||||||||||
Db      1 TGCGCTGTTTTTTCTGGTTCTGCAAGGCAACTGACCTTT 39


RESULT   4
HSU27254
LOCUS       HSU27254                   48 bp    mRNA    linear   PRI 10-JUN-1995
DEFINITION  Human isolate M30 T-cell receptor V-alpha 5/J alpha 22 junction
            mRNA, partial cds.
ACCESSION   U27254
VERSION     U27254.1  GI:857594
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
  TITLE     Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
  JOURNAL   Hum. Immunol. 37 (3), 178-184 (1993)
  MEDLINE   94064390
REFERENCE   2  (bases 1 to 48)
  AUTHORS   Hurwitz,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1995) Julia L. Hurwitz, Immunology, St. Jude
            Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38101,
            USA
FEATURES             Location/Qualifiers
```

```
     source          1. .48
                     /organism="Homo sapiens"
                     /isolate="M30"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
     CDS             <1. .>48
                     /note="encodes V alpha 5/J alpha 22 junction"
                     /codon_start=2
                     /pseudo
                     /evidence=experimental
                     /product="T-cell receptor"
BASE COUNT        7 a     15 c     10 g     16 t
ORIGIN
```

```
  Query Match             77.9%;  Score 30.4;  DB 9;  Length 48;
  Best Local Similarity   96.9%;  Pred. No. 0.012;
  Matches   31;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 TCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           |||| ||||||||||||||||||||||||||||
Db      17 TCCTATCTGGTTCTGCAAGGCAACTGACCTTT 48
```

```
RESULT   5
AF333590
LOCUS       AF333590                 222 bp    mRNA    linear   PRI 02-NOV-2001
DEFINITION  Homo sapiens clone 152.5 T cell receptor alpha chain mRNA, partial
            cds.
ACCESSION   AF333590
VERSION     AF333590.1  GI:13492195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   Talken,B.L., Bailey,C.W., Reardon,S.L., Caldwell,C.W. and
            Hoffman,R.W.
  TITLE     Structural analysis of TCRalpha and beta chains from human T-Cell
            clones specific for small nuclear ribonucleoprotein polypeptides
            Sm-D, Sm-B and U1-70 kDa: TCR complementarity determining region 3
            usage appears highly conserved
  JOURNAL   Scand. J. Immunol. 54 (1-2), 204-210 (2001)
  MEDLINE   21332912
   PUBMED   11439168
REFERENCE   2  (bases 1 to 222)
  AUTHORS   Talken,B.L., Bailey,C.W., Schafermeyer,K.R. and Hoffman,R.W.
  TITLE     T cell receptor alpha and beta chain usage among anti-snRNP
            reactive human T cell clones and lines
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 222)
  AUTHORS   Talken,B.L., Bailey,C.W., Schafermeyer,K.R. and Hoffman,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Internal Medicine, University of Missouri,
            One Hospital Drive, Columbia, MO 65212, USA
FEATURES             Location/Qualifiers
     source          1. .222
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q11-q12"
                     /clone="152.5"
                     /note="isolated from connective tissue disease patient"
     CDS             <1. .>222
                     /codon_start=1
                     /product="T cell receptor alpha chain"
                     /protein_id="AAK28123.1"
                     /db_xref="GI:13492196"
                     /translation="SAMYFCATRGAFSGSARQLTFGSGTQLTVLPDIQNPDPAVYQLR
                     DSKSSDKSVCLFTDFDSQTNVSQSKNSDVY"
     V_segment       <1. .21
                     /note="AV6S1A2N2"
     J_segment       23. .93
                     /note="AJ22"
     C_region        94. .>222
                     /note="AC"
BASE COUNT       58 a     51 c     47 g     66 t
ORIGIN
```

```
  Query Match             76.4%;  Score 29.8;  DB 9;  Length 222;
  Best Local Similarity   93.9%;  Pred. No. 0.022;
  Matches   31;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | | ||||||||||||||||||||||||||||||
Db      31 GCCTTTTCTGGTTCTGCAAGGCAACTGACCTTT 63
```

```
RESULT   6
AF043888
LOCUS       AF043888                 51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV228 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043888
VERSION     AF043888.1  GI:3859395
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV228"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1 [TCRAV1S3]
                     [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
                     and the 5' end of the TCRAJ/TCRBJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72683.1"
                     /db_xref="GI:3859396"
                     /translation="YFCAVRISGSARQLTFG"
BASE COUNT        9 a     10 c     14 g     18 t
ORIGIN
```

```
  Query Match             74.4%;  Score 29;  DB 9;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 0.049;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 TTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           |||||||||||||||||||||||||||||
Db      20 TTTCTGGTTCTGCAAGGCAACTGACCTTT 48
```

```
RESULT   7
```

```
HUMTCAZA
LOCUS       HUMTCAZA                183 bp    mRNA    linear   PRI 13-JAN-1995
DEFINITION  Human T-cell receptor active alpha-chain V-region (V-J-C) mRNA,
            partial cds, clone AA27.
ACCESSION   M17666 J02992
VERSION     M17666.1  GI:338805
KEYWORDS    C-region; J-region; T-cell receptor; V-region; antigen receptor;
            processed gene.
SOURCE      Human peripheral blood lymphocyte, cDNA to mRNA, clone AA27.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183)
  AUTHORS   Klein,M.H., Concannon,P., Everett,M., Kim,L.D., Hunkapiller,T. and
            Hood,L.
  TITLE     Diversity and structure of human T-cell receptor alpha-chain
            variable region genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (19), 6884-6888 (1987)
  MEDLINE   88016194
FEATURES             Location/Qualifiers
     source          1. .183
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
     gene            1. .183
                     /gene="TCRA"
     CDS             <1. .>183
                     /gene="TCRA"
                     /note="T-cell receptor alpha-chain V-region (V-J-C)"
                     /codon_start=3
                     /protein_id="AAA60645.1"
                     /db_xref="GI:553669"
                     /db_xref="GDB:G00-120-404"
                     /translation="VKGINGFEAEFKKSETSFHLTKPSAHMSDAAEYFCAVTFSGSAR
                     QLTFGSGTQLTVLPDI"
     mat_peptide     <1. .>183
                     /gene="TCRA"
                     /product="T-cell receptor alpha-chain"
     misc_recomb     114. .115
                     /gene="TCRA"
                     /organism="Homo sapiens"
     misc_recomb     177. .178
                     /gene="TCRA"
                     /organism="Homo sapiens"
BASE COUNT       44 a     43 c     43 g     53 t
ORIGIN      Chromosome 14q11.2.
```

```
  Query Match           74.4%;  Score 29;  DB 9;  Length 183;
  Best Local Similarity  100.0%;  Pred. No. 0.051;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 TTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           |||||||||||||||||||||||||||||
Db     118 TTTCTGGTTCTGCAAGGCAACTGACCTTT 146
```

```
RESULT   8
AF043877
LOCUS       AF043877                51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV4 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043877
VERSION     AF043877.1  GI:3859383
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV4"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1 [TCRAV1S3]
                     [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV, the n, and the 5'
                     end of the TCRAJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72677.1"
                     /db_xref="GI:3859384"
                     /translation="YFCAVRPSGSARQLTFG"
BASE COUNT        9 a     12 c     13 g     17 t
ORIGIN
```

```
  Query Match           72.8%;  Score 28.4;  DB 9;  Length 51;
  Best Local Similarity  96.7%;  Pred. No. 0.091;
  Matches   29;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      10 CTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | ||||||||||||||||||||||||||||
Db      19 CCTTCTGGTTCTGCAAGGCAACTGACCTTT 48
```

```
RESULT   9
AF043886
LOCUS       AF043886                51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV314 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043886
VERSION     AF043886.1  GI:3859391
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV314"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1 [TCRAV16S1]
                     [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
```

```
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
                     and the 5' end of the TCRAJ/TCRBJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72681.1"
                     /db_xref="GI:3859392"
                     /translation="YFCAVRLTGSARQLTFG"
BASE COUNT       10 a     11 c     13 g     17 t
ORIGIN


  Query Match             72.8%;  Score 28.4;  DB 9;  Length 51;
  Best Local Similarity   96.7%;  Pred. No. 0.091;
  Matches   29;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      10 CTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||| ||||||||||||||||||||||||||
Db      19 CTTACTGGTTCTGCAAGGCAACTGACCTTT 48


RESULT  10
AF043890
LOCUS       AF043890                 51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV109 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043890
VERSION     AF043890.1  GI:3859399
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV109"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1 [TCRAV1S3]
                     [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
                     and the 5' end of the TCRAJ/TCRBJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72685.1"
                     /db_xref="GI:3859400"
                     /translation="YFCAVRRSGSARQLTFG"
BASE COUNT        8 a     11 c     15 g     17 t
ORIGIN


  Query Match             72.8%;  Score 28.4;  DB 9;  Length 51;
  Best Local Similarity   96.7%;  Pred. No. 0.091;
  Matches   29;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      10 CTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | ||||||||||||||||||||||||||||
Db      19 CGTTCTGGTTCTGCAAGGCAACTGACCTTT 48


RESULT  11
HSU30448
LOCUS       HSU30448                 39 bp    mRNA    linear   PRI 01-AUG-1995
DEFINITION  Human isolate M94 T-cell receptor alpha V-J junction (TCR Valpha
            3/J alpha 22) mRNA, partial cds.
ACCESSION   U30448
VERSION     U30448.1  GI:915515
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
  TITLE     Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
  JOURNAL   Hum. Immunol. 37 (3), 178-184 (1993)
  MEDLINE   94064390
REFERENCE   2  (bases 1 to 39)
  AUTHORS   Hurwitz,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="Homo sapiens"
                     /isolate="M94"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
     gene            1. .39
                     /gene="TCR Valpha 3/J alpha 22"
     CDS             <1. .>39
                     /gene="TCR Valpha 3/J alpha 22"
                     /codon_start=1
                     /product="T-cell receptor alpha V-J junction"
                     /protein_id="AAA73621.1"
                     /db_xref="GI:915516"
                     /translation="FCVSSGSARQLTF"
BASE COUNT        5 a     11 c      8 g     15 t
ORIGIN


  Query Match             72.3%;  Score 28.2;  DB 9;  Length 39;
  Best Local Similarity   90.9%;  Pred. No. 0.11;
  Matches   30;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
             |||   ||||||||||||||||||||||||||
Db       7 GTCTCCTCTGGTTCTGCAAGGCAACTGACCTTT 39


RESULT  12
HSU30390
LOCUS       HSU30390                 63 bp    mRNA    linear   PRI 01-AUG-1995
DEFINITION  Human isolate M101 T-cell receptor alpha V-J junction (TCR Valpha
            5/J alpha 22) mRNA, partial cds.
ACCESSION   U30390
VERSION     U30390.1  GI:915415
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63)
  AUTHORS   Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
  TITLE     Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
```

```
  JOURNAL   Hum. Immunol. 37 (3), 178-184 (1993)
  MEDLINE   94064390
REFERENCE   2  (bases 1 to 63)
  AUTHORS   Hurwitz,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES             Location/Qualifiers
     source          1. .63
                     /organism="Homo sapiens"
                     /isolate="M101"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
     gene            1. .63
                     /gene="TCR Valpha 5/J alpha 22"
     CDS             <1. .>63
                     /gene="TCR Valpha 5/J alpha 22"
                     /codon_start=1
                     /product="T-cell receptor alpha V-J junction"
                     /protein_id="AAA73579.1"
                     /db_xref="GI:915416"
                     /translation="TYLCALDPRTVSSGSARQLTF"
BASE COUNT       11 a     20 c     12 g     20 t
ORIGIN


  Query Match             72.3%;  Score 28.2;  DB 9;  Length 63;
  Best Local Similarity   90.9%;  Pred. No. 0.11;
  Matches   30;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||   ||||||||||||||||||||||||||||
Db      31 GTTTCTTCTGGTTCTGCAAGGCAACTGACCTTT 63


RESULT  13
HSTCRAJ3
LOCUS       HSTCRAJ3                  645 bp    DNA     linear   PRI 11-APR-1995
DEFINITION  Human gene for T-cell receptor alpha chain J region.
ACCESSION   X02886
VERSION     X02886.1  GI:36928
KEYWORDS    joining region; T-cell receptor; T-cell receptor alpha.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Yoshikai,Y., Clark,S.P., Taylor,S., Sohn,U., Wilson,B.I.,
            Minden,M.D. and Mak,T.W.
  TITLE     Organization and sequences of the variable, joining and constant
            region genes of the human T-cell receptor alpha-chain
  JOURNAL   Nature 316 (6031), 837-840 (1985)
  MEDLINE   85296333
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    377. .385
                     /note="nonamer recombination signal"
     misc_feature    398. .404
                     /note="heptamer recombination signal"
     J_segment       405. .467
                     /note="J segment (405 is 2nd base in codon) (467 is 1st
                     base in codon)"
BASE COUNT       201 a    107 c    149 g    188 t
ORIGIN


  Query Match             72.3%;  Score 28.2;  DB 9;  Length 645;
  Best Local Similarity   90.9%;  Pred. No. 0.12;
  Matches   30;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||   ||||||||||||||||||||||||||||
Db     404 GTTTCTTCTGGTTCTGCAAGGCAACTGACCTTT 436


RESULT  14
HUAE000662
LOCUS       HUAE000662               71153 bp    DNA     linear   PRI 15-JUN-2001
DEFINITION  Homo sapiens T-cell receptor alpha delta locus from bases 1000498
            to 1071650 (section 5 of 5) of the Complete Nucleotide Sequence.
ACCESSION   AE000662 AE000521
VERSION     AE000662.1  GI:2358068
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71153)
  AUTHORS   Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
            Howard,S., Shan,W., Deshpande,P. and Hood,L.
  TITLE     The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
            organization, sequence, and evolution of 97.6 kb of DNA
  JOURNAL   Genomics 19 (3), 478-493 (1994)
  MEDLINE   94245236
  REMARK    This citation covers from bases 966383-1064019
REFERENCE   2  (bases 1 to 71153)
  AUTHORS   Boysen,C., Simon,M.I. and Hood,L.
  TITLE     Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
            bacterial artificial chromosome clones
  JOURNAL   Genome Res. 7 (4), 330-338 (1997)
  MEDLINE   97264339
   PUBMED   9110172
REFERENCE   3  (bases 1 to 71153)
  AUTHORS   Boysen,C., Inyoul,L., Smith,T.M., Smit,A., Wang,K., Rowen,L. and
            Hood,L.
  TITLE     T-Cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
  JOURNAL   Unpublished
  REMARK    This citation covers bases 1-983545 and bases 1064020-1071650
REFERENCE   4  (bases 1 to 71153)
  AUTHORS   Boysen,C., Lee,I., Smith,T.M., Smit,A., Kai,W., Lee,R. and Leroy,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1997) Department of Molecular Biotechnology,
            University of Washington, Box 357730, Seattle, Washington 98195,
            USA
  REMARK    Complete nucleotide sequence of the human T-cell receptor alpha
            delta locus
COMMENT     Sequencing method: high redundancy shotgun.  Interspersed Repeats
            were identified with RepeatMasker (available from
            http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
            sequence repeats were identified with sputnik (available from
            http://serac.mbt.washington.edu/~chrisa/software/sputnik.html).
FEATURES             Location/Qualifiers
     source          1. .71153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   2362. .2660
                     /rpt_family="AluSc"
     repeat_region   4274. .4574
                     /rpt_family="AluSx"
     repeat_region   9569. .9796
                     /rpt_family="MIR"
     repeat_region   10535. .10834
                     /rpt_family="AluY"
     repeat_region   10835. .11138
                     /rpt_family="AluJb"
     repeat_region   complement(13900. .14190)
                     /rpt_family="AluSq"
     repeat_region   complement(20506. .20810)
                     /rpt_family="AluY"
     repeat_region   20928. .21223
                     /rpt_family="AluSx"
     repeat_region   23083. .23335
                     /rpt_family="MIR"
```

```
     repeat_region    25320. .25618
                      /rpt_family="AluSq"
     repeat_region    complement(27167. .27240)
                      /rpt_family="MIR"
     repeat_region    27306. .27428
                      /rpt_family="MIR2"
     repeat_region    complement(30158. .30403)
                      /rpt_family="MIR"
     repeat_region    33727. .33878
                      /rpt_family="MIR"
     repeat_region    complement(33920. .34213)
                      /rpt_family="AluYa5"
     repeat_region    complement(35275. .35553)
                      /rpt_family="MIR2"
     repeat_region    complement(36848. .36886)
                      /rpt_family="L1MA3"
     repeat_region    39321. .39398
                      /rpt_family="MIR"
     repeat_region    40206. .40261
                      /rpt_family="MIR"
     repeat_region    complement(40433. .40659)
                      /rpt_family="MIR"
     repeat_region    complement(40683. .40848)
                      /rpt_family="AluSx"
     repeat_region    complement(41322. .41444)
                      /rpt_family="MIR"
     repeat_region    complement(45732. .45964)
                      /rpt_family="MIR"
     repeat_region    46927. .46992
                      /rpt_family="MIR2"
     repeat_region    complement(47281. .47388)
                      /rpt_family="MIR2"
     repeat_region    49066. .49073
                      /rpt_type=tandem
                      /rpt_unit=TAT
     repeat_region    49734. .50032
                      /rpt_family="AluY"
     repeat_region    51737. .51940
                      /rpt_family="MER3"
     repeat_region    complement(52892. .53047)
                      /rpt_family="MIR"
     repeat_region    54558. .54732
                      /rpt_family="MIR2"
     repeat_region    complement(54749. .55052)
                      /rpt_family="AluSq"
     repeat_region    55059. .55302
                      /rpt_family="MIR2"
     repeat_region    57276. .57570
                      /rpt_family="AluJo"
     repeat_region    59210. .59460
                      /rpt_family="MER46"
     repeat_region    complement(61706. .61769)
                      /rpt_family="MIR2"
     repeat_region    63506. .63580
                      /rpt_family="MIR"
     repeat_region    63695. .63862
                      /rpt_family="L1ME3"
     repeat_region    63931. .63951
                      /rpt_type=tandem
                      /rpt_unit=AC
     repeat_region    64131. .64281
                      /rpt_family="MIR"
     repeat_region    65845. .66137
                      /rpt_family="AluSx"
     repeat_region    complement(66237. .66480)
                      /rpt_family="MIR2"
     repeat_region    complement(68187. .68457)
                      /rpt_family="AluJo"
     repeat_region    complement(68480. .68777)
                      /rpt_family="AluSq"
     repeat_region    complement(68778. .68837)
                      /rpt_family="FAM"
     repeat_region    69326. .69628
                      /rpt_family="AluSp"
     repeat_region    complement(69688. .69766)
                      /rpt_family="MER58A"
     repeat_region    69845. .69897
                      /rpt_family="MER5A"
     repeat_region    complement(69907. .70048)
                      /rpt_family="MER5A"
     repeat_region    complement(70060. .70225)
                      /rpt_family="MER5B"
BASE COUNT    20125 a  15457 c  16410 g  19161 t
ORIGIN


  Query Match             72.3%;  Score 28.2;  DB 9;  Length 71153;
  Best Local Similarity   90.9%;  Pred. No. 0.13;
  Matches   30;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||   ||||||||||||||||||||||||||||
Db   28162 GTTTCTTCTGGTTCTGCAAGGCAACTGACCTTT 28194


RESULT  15
HUMTCRADCV
LOCUS       HUMTCRADCV              97630 bp    DNA     linear   PRI 25-OCT-2001
DEFINITION  Human Tcr-C-delta gene, exons 1-4; Tcr-V-delta gene, exons 1-2;
            T-cell receptor alpha (Tcr-alpha) gene, J1-J61 segments; and
            Tcr-C-alpha gene, exons 1-4.
ACCESSION   M94081
VERSION     M94081.1  GI:2627263
KEYWORDS    T-cell receptor C-alpha; T-cell receptor C-delta; T-cell receptor
            V-delta; T-cell receptor alpha.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97630)
  AUTHORS   Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
            Howard,S., Shan,W., Deshpande,P. and Hood,L.
  TITLE     The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
            organization, sequence, and evolution of 97.6 kb of DNA
  JOURNAL   Genomics 19 (3), 478-493 (1994)
  MEDLINE   94245236
   PUBMED   8188290
COMMENT     On Nov 20, 1997 this sequence version replaced gi:339242.
FEATURES             Location/Qualifiers
     source          1. .97630
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /germline
     exon            3168. .3464
                     /gene="Tcr-C-delta"
                     /number=1
     gene            3168. .6809
                     /gene="Tcr-C-delta"
     exon            4028. .4098
                     /gene="Tcr-C-delta"
                     /number=2
     exon            4458. .4572
                     /gene="Tcr-C-delta"
                     /number=3
     exon            5787. .6809
                     /gene="Tcr-C-delta"
                     /number=4
     gene            9295. .9868
                     /gene="Tcr-V-delta"
     exon            9295. .9598
                     /gene="Tcr-V-delta"
                     /number=1
     exon            9832. .9868
                     /gene="Tcr-V-delta"
                     /number=2
     gene            <15555. .>85296
```

/gene="Tcr-alpha"
CDS <15555. .>15627
/gene="Tcr-alpha"
/standard_name="possible J 61 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86739.1"
/db_xref="GI:553731"
/translation="SCGYRVNRKLTFGANTRGIMKLS"
exon 16556. .16610
/gene="Tcr-alpha"
/note="putative; possible J 60 gene segment"
CDS <16801. .>16850
/gene="Tcr-alpha"
/standard_name="possible J 59 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86740.1"
/db_xref="GI:553732"
/translation="EGRKQEIYIWNGDASE"
exon 17953. .18016
/gene="Tcr-alpha"
/note="putative; possible J 58 gene segment"
CDS <19121. .>19181
/gene="Tcr-alpha"
/standard_name="possible J 57 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86741.1"
/db_xref="GI:553733"
/translation="TQGGSEKLVFGKGTKLTVNP"
CDS <19765. .>19829
/gene="Tcr-alpha"
/standard_name="possible J 56 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86742.1"
/db_xref="GI:553734"
/translation="CYTGANSKLTFGKGITLSVRP"
CDS <22523. .>22593
/gene="Tcr-alpha"
/standard_name="possible J 55 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86743.1"
/db_xref="GI:553735"
/translation="VLWCNSGSPEAGIWPRNQADYQP"
CDS <23251. .>23316
/gene="Tcr-alpha"
/standard_name="possible J 54 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86744.1"
/db_xref="GI:553736"
/translation="NSGGSNYKLTFGKGTLLTVNP"
CDS <26473. .>26542
/gene="Tcr-alpha"
/standard_name="possible J 53 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86745.1"
/db_xref="GI:553737"
/translation="ANAGGTSYGKLTFGQGTILTVHP"
exon 27429. .27491
/gene="Tcr-alpha"
/note="putative; possible J 52 gene segment"
CDS <27841. .>27905
/gene="Tcr-alpha"
/standard_name="possible J 51 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86746.1"
/db_xref="GI:339250"
/translation="KSLYFWSSCPCAGIVGQADST"
CDS <28839. .>28895
/gene="Tcr-alpha"
/standard_name="possible J 50 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86747.1"
/db_xref="GI:339251"
/translation="ENLLRQGDIWARDKLISHS"
CDS <29734. .>29789
/gene="Tcr-alpha"
/standard_name="possible J 49 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86748.1"
/db_xref="GI:553738"
/translation="NTGNQFYFGTGTSLTVIP"
CDS <30737. .>30799
/gene="Tcr-alpha"
/standard_name="possible J 48 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86749.1"
/db_xref="GI:553739"
/translation="SNFGNEKLTFGTGTRLTIIP"
CDS <33096. .>33152
/gene="Tcr-alpha"
/standard_name="possible J 47 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86750.1"
/db_xref="GI:339254"
/translation="GIWKQTGLWRRNHSESQVL"
CDS <33647. .>33709
/gene="Tcr-alpha"
/standard_name="possible J 46 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86751.1"
/db_xref="GI:553740"
/translation="KKSSGDKLTFGTGTRLAVRP"
CDS <34168. .>34234
/gene="Tcr-alpha"
/standard_name="possible J 45 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86752.1"
/db_xref="GI:553741"
/translation="VYSGGGADGLTFGKGTHLIIQP"
CDS <35064. .>35126
/gene="Tcr-alpha"
/standard_name="possible J 44 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86753.1"
/db_xref="GI:553742"
/translation="NTGTASKLTFGTGTRLQVTL"
CDS <36139. .>36207
/gene="Tcr-alpha"
/standard_name="possible J 43 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86754.1"
/db_xref="GI:553743"
/translation="MYYCDNNNDMRFGAGTRLTVKP"
CDS <37130. .>37195
/gene="Tcr-alpha"
/standard_name="possible J 42 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86755.1"
/db_xref="GI:553744"
/translation="NYGGSQGNLIFGKGTKLSVKP"

```
   CDS             <37899. .>37961
                   /gene="Tcr-alpha"
                   /standard_name="possible J 41 gene segment"
                   /note="putative"
                   /codon_start=1

  Query Match           72.3%;  Score 28.2;  DB 9;  Length 97630;
  Best Local Similarity 90.9%;  Pred. No. 0.13;
  Matches   30;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||   ||||||||||||||||||||||||||||
Db   62271 GTTTCTTCTGGTTCTGCAAGGCAACTGACCTTT 62303


Search completed: June  6, 2002, 07:39:13
Job time: 5376 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:54:26 ; Search time 2649.09 Seconds
(without alignments)
198.703 Million cell updates/sec

Title: US-08-881-509-5
Perfect score: 39
Sequence: 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		30.4	77.9	564	12	AQ221133	AQ221133 HS_3223_B
c 2		26.2	67.2	127	10	BJ140929	BJ140929 BJ140929
3		24.6	63.1	609	12	AZ570267	AZ570267 272PvD08
c 4		24.6	63.1	648	10	BJ138927	BJ138927 BJ138927
5		23	59.0	292	9	BB231582	BB231582 BB231582
c 6		22.8	58.5	294	9	AA234451	AA234451 zr74a08.r
7		22.2	56.9	316	9	BB253473	BB253473 BB253473
c 8		22.2	56.9	393	10	BI289485	BI289485 UI-R-DK0-
c 9		22	56.4	157	10	BG993065	BG993065 RC4-HT110
c 10		22	56.4	267	10	BG375218	BG375218 UI-R-CV1-
c 11		22	56.4	315	9	AA955644	AA955644 UI-R-E1-f
c 12		22	56.4	336	10	BG375188	BG375188 UI-R-CV1-
c 13		22	56.4	420	9	AI176100	AI176100 EST219678
c 14		22	56.4	470	10	BI294666	BI294666 UI-R-DK0-
c 15		22	56.4	508	10	BI293376	BI293376 UI-R-DK0-
c 16		22	56.4	564	12	AQ683171	AQ683171 HS_5378_B
c 17		22	56.4	633	9	AI233243	AI233243 EST229931
18		22	56.4	899	12	CNS0407S	AL268561 Tetraodon
19		22	56.4	1101	12	CNS05SV2	AL352343 Tetraodon
c 20		21.6	55.4	470	12	AQ993790	AQ993790 RPCI-23-3
21		21.6	55.4	551	12	AQ491072	AQ491072 RPCI-11-2
22		21.6	55.4	619	12	BH062049	BH062049 RPCI-24-3
23		21.6	55.4	664	12	AZ313505	AZ313505 1M0029I07
c 24		21.6	55.4	667	9	BB639586	BB639586 BB639586
25		21.6	55.4	796	10	BF120330	BF120330 601756068
c 26		21.4	54.9	269	9	AV294476	AV294476 AV294476
27		21.4	54.9	275	9	BB367011	BB367011 BB367011
28		21.4	54.9	279	9	BB360913	BB360913 BB360913
29		21.4	54.9	282	9	BB461910	BB461910 BB461910
30		21.4	54.9	285	9	BB345796	BB345796 BB345796
31		21.4	54.9	291	9	BB079644	BB079644 BB079644
32		21.4	54.9	294	9	BB402658	BB402658 BB402658
33		21.4	54.9	294	9	BB455396	BB455396 BB455396
34		21.4	54.9	296	9	BB533795	BB533795 BB533795
35		21.4	54.9	305	9	BB375951	BB375951 BB375951
36		21.4	54.9	317	9	BB097763	BB097763 BB097763
37		21.4	54.9	358	9	BB813655	BB813655 BB813655
38		21.4	54.9	370	9	BB813418	BB813418 BB813418
39		21.4	54.9	418	9	BB711165	BB711165 BB711165
40		21.4	54.9	426	9	BB820622	BB820622 BB820622
41		21.4	54.9	432	9	BB787881	BB787881 BB787881
42		21.4	54.9	432	9	BB824353	BB824353 BB824353
43		21.4	54.9	449	9	BB820102	BB820102 BB820102
44		21.4	54.9	470	9	BB822403	BB822403 BB822403
45		21.4	54.9	477	9	AI150530	AI150530 qf41f05.x

ALIGNMENTS

RESULT 1
AQ221133
LOCUS AQ221133 564 bp DNA linear GSS 19-SEP-1998
DEFINITION HS_3223_B2_E10_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3223 Col=20 Row=J, DNA sequence.
ACCESSION AQ221133
VERSION AQ221133.1 GI:3634746
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 564)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3223 row: J column: 20
Class: BAC ends
High quality sequence stop: 564.
FEATURES Location/Qualifiers
source 1. .564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3223 Col=20 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

```
BASE COUNT      160 a    102 c    124 g    171 t      7 others
ORIGIN


  Query Match             77.9%;  Score 30.4;  DB 12;  Length 564;
  Best Local Similarity   96.9%;  Pred. No. 0.18;
  Matches   31;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 TCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | ||||||||||||||||||||||||||||||
Db     419 TTCTTTCTGGTTCTGCAAGGCAACTGACCTTT 450


RESULT   2
BJ140929/c
LOCUS       BJ140929                 127 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ140929 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1164a08 3', mRNA sequence.
ACCESSION   BJ140929
VERSION     BJ140929.1  GI:18301095
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 127)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .127
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1164a08"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
BASE COUNT       45 a     18 c     30 g     34 t
ORIGIN


  Query Match             67.2%;  Score 26.2;  DB 10;  Length 127;
  Best Local Similarity   79.5%;  Pred. No. 5.1;
  Matches   31;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           |||||||||||| | |||    |||||||||  |||||
Db      94 TGCCTCGTCCTTTTTTGTTTCCAAAGGCAACTTGCCTTT 56


RESULT   3
AZ570267
LOCUS       AZ570267                 609 bp    DNA     linear   GSS 15-MAY-2001
DEFINITION  272PvD08 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION   AZ570267
VERSION     AZ570267.1  GI:13981180
KEYWORDS    GSS.
SOURCE      malaria parasite P. vivax.
  ORGANISM  Plasmodium vivax
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Plasmodium vivax"
                     /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
                     497-598)"
                     /db_xref="taxon:5855"
                     /clone_lib="Pv MBN #30"
                     /dev_stage="asexual blood forms"
                     /lab_host="Saimiri boliviensis"
                     /note="Vector: pBluescript SK(+) vector DNA, phagemid
                     excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V;
                     Host leukocytes were extracted from P. vivax infected
                     blood using the following methods: first, infected blood
                     was activated by the addition of 0.5 ml of ADP (40mg/ml)
                     per 10 ml blood. Then blood was passed over a column of
                     acid washed 0.1 mm glass beads, then through a Plasmodipur
                     filter, followed by passage through a column of pre-wet
                     Whatman CF11 powder (1:2 ratio volume of blood to CF11),
                     and finally centrifuged through a 50% Percoll density
                     cushion. Purified DNA was digested with mung bean nuclease
                     in the presence of 44% formamide at 50oC as described
                     (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
                     Nucleic Acids Research 16:6883-6896). Digested DNA was
                     blunt-ended using T4 DNA Polymerase and size fractionated
                     over a Sepharose CL-2B column. Fractions in the size range
                     500bp-4kb were ligated into the Eco RV site of pBluescript
                     SK(+), and E. coli XL-10 Gold transformed with the
                     ligation mixture."
BASE COUNT      127 a    163 c    131 g    188 t
ORIGIN


  Query Match             63.1%;  Score 24.6;  DB 12;  Length 609;
  Best Local Similarity   76.9%;  Pred. No. 30;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           ||||||  ||||| |||| ||  |||| || ||||| ||
Db     432 TGCCTCACCCTTTGTGGTACTAGAAGGGAATTGACCCTT 470


RESULT   4
BJ138927/c
LOCUS       BJ138927                 648 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ138927 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1140h10 3', mRNA sequence.
ACCESSION   BJ138927
VERSION     BJ138927.1  GI:18299093
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
```

Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

```
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1140h10"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
BASE COUNT      202 a    107 c    145 g    194 t
ORIGIN


  Query Match             63.1%;  Score 24.6;  DB 10;  Length 648;
  Best Local Similarity   76.9%;  Pred. No. 31;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
          ||||||||  ||||  |  |||    |||||||||  |||||
Db    639 TGCCTCGTTCTTTTTTGTTTCCAAAGGCAACTTGCCTTT 601


RESULT   5
BB231582
LOCUS       BB231582                292 bp    mRNA    linear   EST 04-JUL-2000
DEFINITION  BB231582 RIKEN full-length enriched, 3 days neonate thymus Mus
            musculus cDNA clone A630033E12 3', mRNA sequence.
ACCESSION   BB231582
VERSION     BB231582.1  GI:8911437
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A630033E12"
                     /clone_lib="RIKEN full-length enriched, 3 days neonate
                     thymus"
                     /tissue_type="thymus"
                     /dev_stage="3 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       69 a     91 c     52 g     80 t
ORIGIN


  Query Match             59.0%;  Score 23;  DB 9;  Length 292;
  Best Local Similarity   74.4%;  Pred. No. 1e+02;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
          | || | ||| ||||| ||| ||| || |  ||||||||
Db    115 TCCCCCATCCATTCTGCTTCGGCAGGGGACTTGACCTTT 153


RESULT   6
AA234451/c
LOCUS       AA234451                294 bp    mRNA    linear   EST 03-MAR-1997
DEFINITION  zr74a08.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:669110 5'
            similar to WP:R90.1 CE06325 PROTEIN KINASE ;, mRNA sequence.
ACCESSION   AA234451
VERSION     AA234451.1  GI:1858970
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
```

High quality sequence stop: 282.

```
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Homo sapiens"
                     /db_xref="GDB:5563077"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:669110"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       76 a     59 c     79 g     80 t
ORIGIN
```

```
  Query Match             58.5%;  Score 22.8;  DB 9;  Length 294;
  Best Local Similarity   79.4%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACC 36
           ||| ||||||| |  |||| || |||| ||||||
Db     283 CCTTGTCCTTTATTTTTCTCCAGGGCAGCTGACC 250
```

```
RESULT   7
BB253473
LOCUS       BB253473                316 bp    mRNA    linear   EST 06-JUL-2000
DEFINITION  BB253473 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730057K23 3' similar to L03305 Mus musculus
            core-binding factor, mRNA sequence.
ACCESSION   BB253473
VERSION     BB253473.1  GI:8946219
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A730057K23"
                     /clone_lib="RIKEN full-length enriched, 7 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="7 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       82 a     64 c     44 g     125 t      1 others
ORIGIN
```

```
  Query Match             56.9%;  Score 22.2;  DB 9;  Length 316;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       5 TCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           || ||||| |||||||||||  ||  ||| || ||
Db      45 TCTTCCTTACTGGTTCTGCACTGCTTCTGTCCGTT 79
```

```
RESULT   8
BI289485/c
LOCUS       BI289485                393 bp    mRNA    linear   EST 19-JUL-2001
DEFINITION  UI-R-DK0-cfh-f-04-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
            UI-R-DK0-cfh-f-04-0-UI 3', mRNA sequence.
ACCESSION   BI289485
VERSION     BI289485.1  GI:14947104
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
```

```
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat aorta pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .393
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-DK0-cfh-f-04-0-UI"
                      /clone_lib="UI-R-DK0"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                      library is a subtracted library derived from a mixture of
                      five individually tagged normalized rat libraries:
                      brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                      aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                      library was constructed from a mixture of equal amounts of
                      RNA from seven different developmental time-points:
                      embryonic day 17, embryonic day 19, embryonic day 21,
                      adult day 1, adult day 12, adult day 75, and adult day
                      200. (Exception: the aorta pool does not contain embryonic
                      day 17 RNA and the placenta pool contains only the three
                      embryonic stages). Each library was normalized
                      individually according to the procedure described by
                      Bonaldo, Lennon & Soares (Genome Research Genome 6:
                      791-806, 1996). For construction of the DK0 subtracted
                      library, plasmid DNA from each of the five individually
                      tagged normalized libraries was mixed in the proportions
                      specified above and electroporated into competent bacteria
                      for production of single-stranded circular DNA
                      representing the pool of libraries. Single-stranded
                      circular DNA representing these five normalized libraries
                      was then used as a tracer in a subtractive hybridization
                      with a driver (PCR amplified inserts from a plasmid DNA
                      template preparation) comprising: a) a set of about 1,000
                      arrayed clones from each of the five non-normalized
                      libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                      aorta (CW0s), and placenta (CX0s). The resulting pool of
                      approximately 5,000 clones represented about 33.3% of the
                      final driver population. A set of about 2,000 arrayed
                      clones from each of the five normalized libraries of brain
                      (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                      placenta (CX0). The resulting pool of about 10,000 clones
                      represented about 66.6% of the final driver population.
                      TAG_LIB=UI-R-DK0
                      TAG_TISSUE=rat aorta pool
                      TAG_SEQ=CTGTAGGATC"
BASE COUNT       119 a     81 c     90 g    103 t
ORIGIN
```

```
  Query Match             56.9%;  Score 22.2;  DB 10;  Length 393;
  Best Local Similarity   77.1%;  Pred. No. 2.3e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 CTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           |||||||||  ||| ||||||  | ||  || ||||
Db     188 CTCGTCCTTATTGGCTCTGCAGGTCACTTGGCCTT 154
```

```
RESULT   9
BG993065/c
LOCUS       BG993065                157 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  RC4-HT1109-060201-013-c07 HT1109 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG993065
VERSION     BG993065.1  GI:14397135
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC4&t2=RC4-HT1109-
            060201-013-c07&t3=2001-02-06&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 25
            High quality sequence stop: 139.
FEATURES              Location/Qualifiers
     source           1. .157
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="HT1109"
                      /dev_stage="Adult"
                      /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       43 a     37 c     45 g     32 t
ORIGIN
```

```
  Query Match             56.4%;  Score 22;  DB 10;  Length 157;
  Best Local Similarity   73.7%;  Pred. No. 2.2e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
             |||  |  ||||  |||  |   |||||||||  ||| ||||
Db      39 TGCATACTCCTGTCTCGAAGTGCAAGGCATCTGCCCTT 2
```

```
RESULT  10
BG375218/c
LOCUS       BG375218                267 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  UI-R-CV1-bsu-f-09-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone
            UI-R-CV1-bsu-f-09-0-UI 3', mRNA sequence.
ACCESSION   BG375218
VERSION     BG375218.1  GI:13299690
```

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 267)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .267
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CV1-bsu-f-09-0-UI"
/clone_lib="UI-R-CV1"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CV1 library is a normalized library constructed from rat eye tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-CV1
TAG_TISSUE=rat eye
TAG_SEQ=CAGCC"

BASE COUNT 65 a 75 c 51 g 76 t
ORIGIN

Query Match 56.4%; Score 22; DB 10; Length 267;
Best Local Similarity 73.7%; Pred. No. 2.5e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
          ||| || ||||| |||||||||   ||| | ||  |||
Db    203 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 166
```

RESULT 11
AA955644/c
LOCUS AA955644 315 bp mRNA linear EST 04-JUL-1999
DEFINITION UI-R-E1-fc-e-04-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone UI-R-E1-fc-e-04-0-UI 3', mRNA sequence.
ACCESSION AA955644
VERSION AA955644.1 GI:4238039
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 315)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On May 7, 1998 this sequence version replaced gi:3119339.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 12-Day-Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1771911
Seq primer: M13 Forward
POLYA=No.

FEATURES Location/Qualifiers
source 1. .315
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E1-fc-e-04-0-UI"
/clone_lib="UI-R-E1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1 library is a subtracted library derived from the UI-R-E0 library. The UI-R-E0 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-E1) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-E0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-E0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-E1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 78 a 93 c 63 g 81 t
ORIGIN

Query Match 56.4%; Score 22; DB 9; Length 315;
Best Local Similarity 73.7%; Pred. No. 2.6e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
          ||| || ||||| |||||||||   ||| | ||  |||
Db    213 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 176
```

RESULT 12
BG375188/c
LOCUS BG375188 336 bp mRNA linear EST 12-MAR-2001
DEFINITION UI-R-CV1-bsu-b-09-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone UI-R-CV1-bsu-b-09-0-UI 3', mRNA sequence.

ACCESSION BG375188
VERSION BG375188.1 GI:13299660
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 336)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .336
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CV1-bsu-b-09-0-UI"
/clone_lib="UI-R-CV1"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CV1 library is a normalized library constructed from rat eye tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-CV1
TAG_TISSUE=rat eye
TAG_SEQ=CAGCC"
BASE COUNT 80 a 100 c 71 g 85 t
ORIGIN

Query Match 56.4%; Score 22; DB 10; Length 336;
Best Local Similarity 73.7%; Pred. No. 2.6e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           ||| || ||||| |||||||||   ||| | ||  |||
Db     212 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 175
```

RESULT 13
AI176100/c
LOCUS AI176100 420 bp mRNA linear EST 20-JAN-1999
DEFINITION EST219678 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone ROVBK59 3' end, mRNA sequence.
ACCESSION AI176100
VERSION AI176100.1 GI:3726738
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 420)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: TC48415
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .420
/organism="Rattus sp."
/db_xref="ATCC (inhost):2031158"
/db_xref="taxon:10118"
/clone="ROVBK59"
/clone_lib="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 105 a 130 c 98 g 87 t
ORIGIN

Query Match 56.4%; Score 22; DB 9; Length 420;
Best Local Similarity 73.7%; Pred. No. 2.8e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           ||| || ||||| |||||||||   ||| | ||  |||
Db     194 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 157
```

RESULT 14
BI294666/c
LOCUS BI294666 470 bp mRNA linear EST 19-JUL-2001
DEFINITION UI-R-DK0-cen-e-05-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone UI-R-DK0-cen-e-05-0-UI 3', mRNA sequence.
ACCESSION BI294666
VERSION BI294666.1 GI:14957341
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 470)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat brain pool library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through

```
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-DK0-cen-e-05-0-UI"
                      /clone_lib="UI-R-DK0"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                      library is a subtracted library derived from a mixture of
                      five individually tagged normalized rat libraries:
                      brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                      aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                      library was constructed from a mixture of equal amounts of
                      RNA from seven different developmental time-points:
                      embryonic day 17, embryonic day 19, embryonic day 21,
                      adult day 1, adult day 12, adult day 75, and adult day
                      200. (Exception: the aorta pool does not contain embryonic
                      day 17 RNA and the placenta pool contains only the three
                      embryonic stages). Each library was normalized
                      individually according to the procedure described by
                      Bonaldo, Lennon & Soares (Genome Research Genome 6:
                      791-806, 1996). For construction of the DK0 subtracted
                      library, plasmid DNA from each of the five individually
                      tagged normalized libraries was mixed in the proportions
                      specified above and electroporated into competent bacteria
                      for production of single-stranded circular DNA
                      representing the pool of libraries. Single-stranded
                      circular DNA representing these five normalized libraries
                      was then used as a tracer in a subtractive hybridization
                      with a driver (PCR amplified inserts from a plasmid DNA
                      template preparation) comprising: a) a set of about 1,000
                      arrayed clones from each of the five non-normalized
                      libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                      aorta (CW0s), and placenta (CX0s). The resulting pool of
                      approximately 5,000 clones represented about 33.3% of the
                      final driver population. A set of about 2,000 arrayed
                      clones from each of the five normalized libraries of brain
                      (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                      placenta (CX0). The resulting pool of about 10,000 clones
                      represented about 66.6% of the final driver population.
                      TAG_LIB=UI-R-DK0
                      TAG_TISSUE=rat brain pool
                      TAG_SEQ=ACTTC"
BASE COUNT      121 a    140 c    101 g    108 t
ORIGIN


  Query Match             56.4%;  Score 22;  DB 10;  Length 470;
  Best Local Similarity   73.7%;  Pred. No. 2.8e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           ||| || ||||| |||||||||   ||| | ||  |||
Db     212 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 175


RESULT  15
BI293376/c
LOCUS       BI293376                 508 bp    mRNA    linear   EST 19-JUL-2001
DEFINITION  UI-R-DK0-cdo-d-10-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
            UI-R-DK0-cdo-d-10-0-UI 3', mRNA sequence.
ACCESSION   BI293376
VERSION     BI293376.1  GI:14954814
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .508
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-DK0-cdo-d-10-0-UI"
                      /clone_lib="UI-R-DK0"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                      library is a subtracted library derived from a mixture of
                      five individually tagged normalized rat libraries:
                      brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                      aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                      library was constructed from a mixture of equal amounts of
                      RNA from seven different developmental time-points:
                      embryonic day 17, embryonic day 19, embryonic day 21,
                      adult day 1, adult day 12, adult day 75, and adult day
                      200. (Exception: the aorta pool does not contain embryonic
                      day 17 RNA and the placenta pool contains only the three
                      embryonic stages). Each library was normalized
                      individually according to the procedure described by
                      Bonaldo, Lennon & Soares (Genome Research Genome 6:
                      791-806, 1996). For construction of the DK0 subtracted
                      library, plasmid DNA from each of the five individually
                      tagged normalized libraries was mixed in the proportions
                      specified above and electroporated into competent bacteria
                      for production of single-stranded circular DNA
                      representing the pool of libraries. Single-stranded
                      circular DNA representing these five normalized libraries
                      was then used as a tracer in a subtractive hybridization
                      with a driver (PCR amplified inserts from a plasmid DNA
                      template preparation) comprising: a) a set of about 1,000
                      arrayed clones from each of the five non-normalized
                      libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                      aorta (CW0s), and placenta (CX0s). The resulting pool of
                      approximately 5,000 clones represented about 33.3% of the
                      final driver population. A set of about 2,000 arrayed
                      clones from each of the five normalized libraries of brain
                      (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                      placenta (CX0). The resulting pool of about 10,000 clones
                      represented about 66.6% of the final driver population.
                      TAG_SEQ=None found"
BASE COUNT      135 a    159 c    117 g     97 t
ORIGIN


  Query Match             56.4%;  Score 22;  DB 10;  Length 508;
  Best Local Similarity   73.7%;  Pred. No. 2.9e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;
```

```
Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           ||| || ||||| |||||||||   ||| | ||  |||
Db     218 TGCTTCCTCCTTCCTGGTTCTGTGTGGCCAGTGGTCTT 181
```

Search completed: June 6, 2002, 06:54:29
Job time: 2692 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:00:20 ; Search time 333.99 Seconds
(without alignments)
200.484 Million cell updates/sec

Title: US-08-881-509-5
Perfect score: 39
Sequence: 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	39	100.0	39	19	AAV18707	cDNA for T-cell re
	2	27.6	70.8	36	19	AAV18708	cDNA for T-cell re
	3	26.2	67.2	1341	19	AAV18705	cDNA for T-cell re
c	4	22.8	58.5	1200	22	AAS06723	Polynucleotide seq
c	5	22.8	58.5	2452	22	AAF44667	Novel protein kina
c	6	22.8	58.5	3735	22	AAI66827	Human protein kina
c	7	22.8	58.5	5499	22	AAI66826	Human protein kina
	8	21.6	55.4	1052	22	AAH94558	Human foetal cDNA,
c	9	21.4	54.9	1664	23	AAS89146	DNA encoding novel
c	10	21	53.8	10732	21	AAA10594	Gene encoding a su
c	11	20.6	52.8	11992	22	AAI99007	Human excretory re
c	12	20.6	52.8	11992	22	AAK80050	Human immune/haema
c	13	20.6	52.8	11992	22	AAI63357	Human kidney relat
	14	20.4	52.3	1001	21	AAH51555	Human FMO related
	15	20.4	52.3	1378	23	AAS81354	DNA encoding novel
	16	20.4	52.3	4462	21	AAA11283	Rat MAGUIN 2 cDNA
	17	20.4	52.3	26016	19	AAV57272	Human flavin-conta
c	18	20.2	51.8	1049	21	AAA97920	A. thaliana PUP1 D
c	19	20.2	51.8	4494	23	ABL05292	Drosophila melanog
c	20	20	51.3	728	16	AAQ83561	Renal carcinoma as
c	21	20	51.3	465237	24	ABA90193	Human oestrogen re
c	22	19.8	50.8	1000	19	AAV08382	Cellulose synthase
	23	19.8	50.8	2067	22	AAH17966	Human cDNA sequenc
	24	19.8	50.8	2533	22	AAS36156	Human cardiovascul
c	25	19.8	50.8	2967	23	AAS81075	DNA encoding novel
	26	19.8	50.8	13176	22	AAS36157	Human cardiovascul
c	27	19.6	50.3	401	22	AAK95550	Human neuregulin g
c	28	19.6	50.3	401	22	AAK97043	Human neuregulin g
c	29	19.6	50.3	4482	23	AAS72747	DNA encoding novel
c	30	19.6	50.3	4482	23	AAS88068	DNA encoding novel
c	31	19.6	50.3	4974	23	AAS72749	DNA encoding novel
c	32	19.6	50.3	4974	23	AAS88070	DNA encoding novel
	33	19.4	49.7	435	21	AAA51458	PCpB-DFIAEGGGVR-hP
	34	19.4	49.7	1517	23	AAS91336	DNA encoding novel
c	35	19.4	49.7	1586	21	AAF18247	Lung cancer associ
c	36	19.4	49.7	1614	23	AAS81161	DNA encoding novel
c	37	19.4	49.7	3174	23	ABL17583	Drosophila melanog
	38	19.4	49.7	5616	23	ABL17582	Drosophila melanog
c	39	19.2	49.2	393	21	AAC06405	Human secreted pro
c	40	19.2	49.2	397	22	AAL20375	Human breast cance
c	41	19.2	49.2	750	21	AAC77592	Human ORFX ORF3147
	42	19.2	49.2	1636	22	AAI60336	Human polynucleoti
c	43	19.2	49.2	1650	22	AAI58550	Human polynucleoti
	44	19.2	49.2	1788	20	AAX60777	Soybean cytochrome
	45	19.2	49.2	3172	22	AAS34789	cDNA encoding nove

ALIGNMENTS

RESULT 1
AAV18707
ID AAV18707 standard; cDNA; 39 BP.
XX
AC AAV18707;
XX
DT 26-JUN-1998 (first entry)
XX
DE cDNA for T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..39
FT /*tag= a
XX
PN DE19625191-A1.
XX
PD 02-JAN-1998.
XX
PF 24-JUN-1996; 96DE-1025191.
XX
PR 24-JUN-1996; 96DE-1025191.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Schendel D;
XX

```
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47590.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence encodes the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence 39 BP; 5 A; 12 C; 8 G; 14 T; 0 other;

  Query Match             100.0%;  Score 39;  DB 19;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-07;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           |||||||||||||||||||||||||||||||||||||||
Db       1 tgcctcgtcctttctggttctgcaaggcaactgaccttt 39


RESULT   2
AAV18708
ID   AAV18708 standard; cDNA; 36 BP.
XX
AC   AAV18708;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1..36
FT                   /*tag=  a
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47591.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence encodes the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence 36 BP; 6 A; 11 C; 8 G; 11 T; 0 other;

  Query Match             70.8%;  Score 27.6;  DB 19;  Length 36;
  Best Local Similarity   88.2%;  Pred. No. 0.032;
  Matches   30;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       6 CGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | ||  | ||||||||||||||||||||||||||||
Db       3 cctcgctactggttctgcaaggcaactgaccttt 36


RESULT   3
AAV18705
ID   AAV18705 standard; cDNA; 1341 BP.
XX
AC   AAV18705;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor alpha-chain.
XX
KW   Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..804
FT                   /*tag=  a
FT   sig_peptide     1..54
FT                   /*tag=  b
FT   mat_peptide     55..801
FT                   /*tag=  c
FT                   /product= T-cell_receptor_alpha-chain
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47588.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Pages 11-13; 30pp; German.
XX
CC   The present sequence encodes the alpha-chain of a human T-cell
CC   receptor (TCR), which can be used in the diagnosis, monitoring,
CC   prevention and therapy of a tumour disease, specifically renal
CC   cell carcinoma.
XX
SQ   Sequence 1341 BP; 331 A; 365 C; 294 G; 351 T; 0 other;

  Query Match             67.2%;  Score 26.2;  DB 19;  Length 1341;
  Best Local Similarity   79.5%;  Pred. No. 0.27;
  Matches   31;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | |  | || |   |||||||||||||||||||||||||
Db     310 tactgcctcgtgggtggttctgcaaggcaactgaccttt 348


RESULT   4
AAS06723/c
ID   AAS06723 standard; cDNA; 1200 BP.
```

XX
AC AAS06723;
XX
DT 12-SEP-2001 (first entry)
XX
DE Polynucleotide sequence encoding human protein kinase #23.
XX
KW Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200138503-A2.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000WO-US32085.
XX
PR 24-NOV-1999; 99US-0167482.
XX
PA (SUGE-) SUGEN INC.
XX
PI Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI Flanagan P, Clary D;
XX
DR WPI; 2001-343950/36.
DR P-PSDB; AAU03523.
XX
PT Nucleic acids encoding human kinase polypeptides, useful for preventing
PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and
PT neuronal-associated diseases, and microbial infections -
XX
PS Example 1; Figure 1; 433pp; English.
XX
CC AAS06701-AAS06757 encode for novel human protein kinases #1-57. The
CC novel protein kinases have been identified as members of the tyrosine
CC or serine/threonine kinase (PTK and STK) families. The polynucleotides
CC encoding protein kinases and the polypeptides may be used in the
CC prevention, diagnosis and treatment of diseases associated with
CC inappropriate kinase expression. For example, they may be used to treat
CC cancers (especially cancers of haematopoietic origin), cardiovascular
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
CC immune related diseases (e.g. rheumatoid arthritis), neurological
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
CC Additionally, polynucleotides encoding protein kinases may be
CC used for gene therapy and as DNA probes in diagnostic assays.
CC The protein kinase polypeptides may be used as antigens in the production
CC of antibodies against the protein kinases and in assays to identify
CC modulators of protein kinase expression and activity.
XX
SQ Sequence 1200 BP; 333 A; 245 C; 299 G; 320 T; 3 other;

Query Match 58.5%; Score 22.8; DB 22; Length 1200;
Best Local Similarity 79.4%; Pred. No. 7;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACC 36
||| ||||||| | |||| || |||| ||||||
Db 697 CCTTGTCCTTTATTTTTCTCCAGGGCAGCTGACC 664

RESULT 5
AAF44667/c
ID AAF44667 standard; cDNA; 2452 BP.
XX
AC AAF44667;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase cDNA, SEQ ID NO: 47.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGE-) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
DR P-PSDB; AAB65640.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Example 4; Fig 2; 310pp; English.
XX
CC The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 2452 BP; 676 A; 529 C; 624 G; 622 T; 1 other;

Query Match 58.5%; Score 22.8; DB 22; Length 2452;
Best Local Similarity 79.4%; Pred. No. 8.1;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACC 36
||| ||||||| | |||| || |||| ||||||
Db 1102 CCTTGTCCTTTATTTTTCTCCAGGGCAGCTGACC 1069

RESULT 6
AAI66827/c
ID AAI66827 standard; cDNA; 3735 BP.
XX
AC AAI66827;
XX
DT 07-JAN-2002 (first entry)
XX
DE Human protein kinase polypeptide 23546 coding sequence.
XX

```
KW   Protein kinase; 3714; 16742; 23546; 13887; cancer; bone disorder; human;
KW   cytostatic; antiinflammatory; immunosuppressive; cardiant; hepatotrophic;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3755
FT                   /*tag=  a
FT                   /product= "protein kinase 23546"
XX
PN   WO200173050-A2.
XX
PD   04-OCT-2001.
XX
PF   23-MAR-2001; 2001WO-US09483.
XX
PR   24-MAR-2000; 2000US-191846P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers R;
XX
DR   WPI; 2001-611632/70.
DR   P-PSDB; AAG65766.
XX
PT   New human protein kinase polypeptides, 3714, 16742, 23546 and 13887,
PT   useful in diagnosis of cancer or cellular proliferation or
PT   differentiation disorders and to screen for polypeptide modulators
PT   useful to treat such conditions
XX
PS   Claim 1; Fig 7A-G; 169pp; English.
XX
CC   The invention provides novel human protein kinase polypeptides, 3714,
CC   16742, 23546 and 13887 and nucleic acid molecules encoding them. The
CC   protein kinase polypeptides can be expressed by standard recombinant
CC   methodology.3714, 16742, 23546 or 13887 nucleic acids and polypeptides
CC   are useful for diagnostic and screening methods to identify subjects (at
CC   risk of) having cancer or cellular proliferation and/or differentiation
CC   disorders. 3714, 16742, 23546 or 13887 nucleic acids, polypeptides and
CC   modulators are useful for the treatment of cancer, particularly colon
CC   cancer or cellular proliferation and/or differentiation disorders. Other
CC   disorders associated with 3714, 16742, 23546 or 13887 expression or
CC   activity that can be treated include bone related disorders, inflammatory
CC   disorders, autoimmune diseases, cardiovascular disorders and liver
CC   diseases. The present sequence represents a human protein kinase
CC   polypeptide 23546 coding sequence.
XX
SQ   Sequence 3735 BP; 1088 A; 878 C; 883 G; 886 T; 0 other;

  Query Match             58.5%;  Score 22.8;  DB 22;  Length 3735;
  Best Local Similarity   79.4%;  Pred. No. 8.9;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACC 36
           ||| ||||||| |  |||| || |||| ||||||
Db     697 CCTTGTCCTTTATTTTTCTCCAGGGCAGCTGACC 664


RESULT   7
AAI66826/c
ID   AAI66826 standard; cDNA; 5499 BP.
XX
AC   AAI66826;
XX
DT   07-JAN-2002  (first entry)
XX
DE   Human protein kinase polypeptide 23546 encoding cDNA.
XX
KW   Protein kinase; 3714; 16742; 23546; 13887; cancer; bone disorder; human;
KW   cytostatic; antiinflammatory; immunosuppressive; cardiant; hepatotrophic;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..316
FT                   /*tag=  a
FT   CDS             317..4051
FT                   /*tag=  b
FT                   /product= "protein kinase 23546"
FT                   /note= "coding sequence specifically claimed"
FT   3'UTR           4052..5499
FT                   /*tag=  c
XX
PN   WO200173050-A2.
XX
PD   04-OCT-2001.
XX
PF   23-MAR-2001; 2001WO-US09483.
XX
PR   24-MAR-2000; 2000US-191846P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers R;
XX
DR   WPI; 2001-611632/70.
DR   P-PSDB; AAG65766.
XX
PT   New human protein kinase polypeptides, 3714, 16742, 23546 and 13887,
PT   useful in diagnosis of cancer or cellular proliferation or
PT   differentiation disorders and to screen for polypeptide modulators
PT   useful to treat such conditions
XX
PS   Claim 1; Fig 7A-G; 169pp; English.
XX
CC   The invention provides novel human protein kinase polypeptides, 3714,
CC   16742, 23546 and 13887 and nucleic acid molecules encoding them. The
CC   protein kinase polypeptides can be expressed by standard recombinant
CC   methodology.3714, 16742, 23546 or 13887 nucleic acids and polypeptides
CC   are useful for diagnostic and screening methods to identify subjects (at
CC   risk of) having cancer or cellular proliferation and/or differentiation
CC   disorders. 3714, 16742, 23546 or 13887 nucleic acids, polypeptides and
CC   modulators are useful for the treatment of cancer, particularly colon
CC   cancer or cellular proliferation and/or differentiation disorders. Other
CC   disorders associated with 3714, 16742, 23546 or 13887 expression or
CC   activity that can be treated include bone related disorders, inflammatory
CC   disorders, autoimmune diseases, cardiovascular disorders and liver
CC   diseases. The present sequence represents a human protein kinase
CC   polypeptide 23546 encoding cDNA.
XX
SQ   Sequence 5499 BP; 1600 A; 1215 C; 1258 G; 1426 T; 0 other;

  Query Match             58.5%;  Score 22.8;  DB 22;  Length 5499;
  Best Local Similarity   79.4%;  Pred. No. 9.7;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACC 36
           ||| ||||||| |  |||| || |||| ||||||
Db    1013 CCTTGTCCTTTATTTTTCTCCAGGGCAGCTGACC 980


RESULT   8
AAH94558
ID   AAH94558 standard; cDNA; 1052 BP.
XX
AC   AAH94558;
XX
DT   05-OCT-2001  (first entry)
XX
DE   Human foetal cDNA, SEQ ID NO: 1245.
XX
KW   Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
```

KW nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200155339-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02723.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPI; 2001-465571/50.
DR P-PSDB; AAM06883.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Example 3; Page 705; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a full length cDNA which was assembled using expressed
CC sequence tags (ESTs) found to be expressed in human foetal tissue
CC cDNA libraries as seeds.
XX
SQ Sequence 1052 BP; 299 A; 188 C; 248 G; 316 T; 1 other;

Query Match 55.4%; Score 21.6; DB 22; Length 1052;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
Db 407 cttaatcttttctgtttctgctagccatgtgacctt 442

RESULT 9
AAS89146/c
ID AAS89146 standard; cDNA; 1664 BP.
XX
AC AAS89146;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24950.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24959.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24950; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1664 BP; 407 A; 499 C; 425 G; 333 T; 0 other;

Query Match 54.9%; Score 21.4; DB 23; Length 1664;
Best Local Similarity 71.8%; Pred. No. 29;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
Db 1048 TTCCTCGTCATTTCTGGCCCAGCTGGTCCACTGGGCTTT 1010

RESULT 10
AAA10594/c
ID AAA10594 standard; DNA; 10732 BP.
XX
AC AAA10594;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KW Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PF 26-AUG-1998; 98JP-0239998.
XX

```
PR   26-AUG-1998;   98JP-0239998.
XX
PA   (MIZU/) MIZUNO K.
PA   (OJIP ) OJI PAPER CO.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; AAY85179.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;


  Query Match             53.8%;  Score 21;  DB 21;  Length 10732;
  Best Local Similarity   59.0%;  Pred. No. 63;
  Matches   23;  Conservative    7;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           :: :| :| : :|||: |||||||||| | |||| || |
Db    6645 YRASTTBTTYCYTCTBCTTCTGCAAGGGATCTGAACTAT 6607


RESULT  11
AAI99007/c
ID   AAI99007 standard; DNA; 11992 BP.
XX
AC   AAI99007;
XX
DT   07-JAN-2002  (first entry)
XX
DE   Human excretory related polynucleotide SEQ ID NO 771.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant;  immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW   excretory system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155313-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01323.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
```

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465569/50.
XX
PT Isolated nucleic acid molecule encoding excretory system antigen is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Example 2; SEQ ID NO 771; 574pp + Sequence Listing; English.
XX
CC The invention relates to novel excretory system related human
CC polynucleotides (AAI98567-AAI99503) and the encoded proteins
CC (AAM99594-AAM99913) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, espeically
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 11992 BP; 2885 A; 3018 C; 2979 G; 3110 T; 0 other;

Query Match 52.8%; Score 20.6; DB 22; Length 11992;
Best Local Similarity 74.3%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCT 37
||| ||| | |||| || ||| |||||||| ||
Db 6848 CCTTCACCTCTATGGTCCTCCAATGCAACTGAACT 6814

RESULT 12
AAK80050/c
ID AAK80050 standard; DNA; 11992 BP.
XX
AC AAK80050;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34862.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 34862; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 11992 BP; 2885 A; 3018 C; 2979 G; 3110 T; 0 other;

Query Match 52.8%; Score 20.6; DB 22; Length 11992;
Best Local Similarity 74.3%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCT 37
||| ||| | |||| || ||| |||||||| ||
Db 6848 CCTTCACCTCTATGGTCCTCCAATGCAACTGAACT 6814

RESULT 13
AAI63357/c
ID AAI63357 standard; DNA; 11992 BP.
XX
AC AAI63357;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human kidney related polynucleotide SEQ ID NO 672.
XX
KW Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmological; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
OS Homo sapiens.
XX
PN WO200155323-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01343.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488784/53.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Disclosure; SEQ ID NO 672; 564pp + Sequence Listing; English.
XX
CC The invention relates to novel kidney related polynucleotides
CC (AAI62971-AAI63793) and the encoded polypeptides (AAM42417-AAM42691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polynucleotides and proteins are also
CC useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. The genes are isolated from a range
CC of human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 11992 BP; 2885 A; 3018 C; 2979 G; 3110 T; 0 other;

Query Match 52.8%; Score 20.6; DB 22; Length 11992;
Best Local Similarity 74.3%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 CCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCT 37
||| ||| | |||| || ||| |||||||| ||
Db 6848 CCTTCACCTCTATGGTCCTCCAATGCAACTGAACT 6814

RESULT 14
AAH51555
ID AAH51555 standard; DNA; 1001 BP.
XX
AC AAH51555;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human FMO related DNA containing a biallelic polymorphism SEQ ID 446.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 622; 673pp; English.
XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 1001 BP; 263 A; 207 C; 213 G; 318 T; 0 other;

Query Match 52.3%; Score 20.4; DB 21; Length 1001;
Best Local Similarity 71.1%; Pred. No. 68;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
|| | | || | || |||| |||| |||||| ||||
Db 460 tgatgcctttctctgtgtttcttcaagacaactgtcctt 497

RESULT 15
AAS81354
ID AAS81354 standard; cDNA; 1378 BP.
XX
AC AAS81354;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #17158.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG17167.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 17158; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1378 BP; 253 A; 413 C; 356 G; 356 T; 0 other;

Query Match 52.3%; Score 20.4; DB 23; Length 1378;
Best Local Similarity 71.1%; Pred. No. 73;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
           | | ||  ||||||| |||||   |||| ||  |||||
Db     286 ggcccggtctttctgcttctggctggcatcttcccttt 323
```

Search completed: June 6, 2002, 07:00:23
Job time: 3046 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:38:13 ; Search time 78.65 Seconds
(without alignments)
121.802 Million cell updates/sec

Title: US-08-881-509-5
Perfect score: 39
Sequence: 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.2	49.2	1788	3	US-08-948-564-9	Sequence 9, Appli
c	2	19	48.7	11459	4	US-09-454-721A-3	Sequence 3, Appli
	3	18.8	48.2	1713	4	US-09-181-336-12	Sequence 12, Appl
	4	18.8	48.2	1721	2	US-08-922-170B-9	Sequence 9, Appli
	5	18.8	48.2	1721	2	US-08-922-170B-11	Sequence 11, Appl
	6	18.8	48.2	1721	4	US-09-071-739B-1	Sequence 1, Appli
	7	18.8	48.2	1721	4	US-09-071-739B-3	Sequence 3, Appli
	8	18.8	48.2	1721	4	US-09-260-038B-1	Sequence 1, Appli
	9	18.8	48.2	1721	4	US-09-260-038B-3	Sequence 3, Appli
	10	18.8	48.2	1723	4	US-09-181-336-14	Sequence 14, Appl
c	11	18.6	47.7	2325	1	US-08-019-870-2	Sequence 2, Appli
c	12	18.6	47.7	2325	1	US-08-019-870-4	Sequence 4, Appli
c	13	18.6	47.7	2325	1	US-08-019-870-7	Sequence 7, Appli
c	14	18.6	47.7	2325	1	US-08-019-870-10	Sequence 10, Appl
c	15	18.6	47.7	2325	1	US-08-314-309A-20	Sequence 20, Appl
c	16	18.6	47.7	2325	1	US-08-633-760-43	Sequence 43, Appl
c	17	18.6	47.7	2325	1	US-08-633-760-45	Sequence 45, Appl
c	18	18.6	47.7	2325	1	US-08-633-760-47	Sequence 47, Appl
c	19	18.6	47.7	2325	1	US-08-633-760-49	Sequence 49, Appl
c	20	18.6	47.7	2325	1	US-08-633-760-51	Sequence 51, Appl
c	21	18.6	47.7	2373	1	US-08-019-870-9	Sequence 9, Appli
c	22	18.6	47.7	2847	1	US-07-747-901A-2	Sequence 2, Appli
c	23	18.6	47.7	2847	1	US-07-935-312-2	Sequence 2, Appli
	24	18.4	47.2	2220	6	5171850-1	Patent No. 5171850
c	25	18.4	47.2	2585	4	US-09-008-697A-7	Sequence 7, Appli
c	26	18.4	47.2	5496	3	US-08-600-982-23	Sequence 23, Appl
c	27	18.4	47.2	5496	5	PCT-US94-10261A-23	Sequence 23, Appl
	28	18.2	46.7	301	4	US-09-439-313-250	Sequence 250, App
	29	18.2	46.7	420	4	US-08-943-731-198	Sequence 198, App
c	30	18.2	46.7	1383	1	US-08-289-709-2	Sequence 2, Appli
c	31	18.2	46.7	1383	1	US-08-602-656-2	Sequence 2, Appli
	32	18.2	46.7	1474	1	US-08-465-980-1	Sequence 1, Appli
	33	18.2	46.7	1474	2	US-09-053-303-1	Sequence 1, Appli
	34	18.2	46.7	1474	5	PCT-US95-07093-1	Sequence 1, Appli
	35	18.2	46.7	1828	2	US-08-888-497-29	Sequence 29, Appl
	36	18.2	46.7	1828	4	US-09-362-230-29	Sequence 29, Appl
	37	18.2	46.7	1828	5	PCT-US94-07926-29	Sequence 29, Appl
c	38	18.2	46.7	3083	4	US-08-956-657-1	Sequence 1, Appli
	39	18.2	46.7	6642	1	US-08-727-034-5	Sequence 5, Appli
	40	18.2	46.7	6843	1	US-08-727-034-6	Sequence 6, Appli
	41	18.2	46.7	20084	4	US-08-943-731-5	Sequence 5, Appli
c	42	18	46.2	1392	4	US-09-163-444-1	Sequence 1, Appli
c	43	18	46.2	1735	4	US-09-163-444-3	Sequence 3, Appli
c	44	18	46.2	2768	1	US-08-321-478-1	Sequence 1, Appli
c	45	18	46.2	2768	1	US-08-321-478-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-948-564-9
; Sequence 9, Application US/08948564
; Patent No. 6121512
;  GENERAL INFORMATION:
;    APPLICANT:  Siminszky, Balazs
;    APPLICANT:  Dewey, Ralph E.
;    APPLICANT:  Corbin, Frederick T.
;    TITLE OF INVENTION:  No. 6121512el Cytochrome P-450 Constructs and
;    TITLE OF INVENTION:  Methods of Producing Herbicide-Resistant Transgenic Plants
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Virginia C. Bennett
;      STREET:  PO Box 37428
;      CITY:  Raleigh
;      STATE:  No. 6121512th Carolina
;      COUNTRY:  USA
;      ZIP:  27627
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,564
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bennett, Virginia C.
;      REGISTRATION NUMBER:  37,092
;      REFERENCE/DOCKET NUMBER:  5051-409
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-854-1400
;      TELEFAX:  919-854-1401
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1788 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6..1601
US-08-948-564-9


  Query Match           49.2%;  Score 19.2;  DB 3;  Length 1788;
  Best Local Similarity  87.5%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy       4 CTCGTCCTTTCTGGTTCTGCAAGG 27
           ||||||| || |||||||| ||||
Db    1383 CTCGTCCCTTTTGGTTCTGGAAGG 1406


RESULT   2
US-09-454-721A-3/c
; Sequence 3, Application US/09454721A
; Patent No. 6296854
; GENERAL INFORMATION:
;  APPLICANT:    Peter Pushko
;  APPLICANT:  Michael D. Parker
;  APPLICANT:  Jonathan F. Smith
;  APPLICANT:  Burce J. Crise
;  TITLE OF INVENTION:  Live Attenuated Venezuelan Equine Encephalitis Vaccine
;  FILE REFERENCE:  Army 146
;  CURRENT APPLICATION NUMBER: US/09/454,721A
;  CURRENT FILING DATE:  1999-12-07
;  PRIOR APPLICATION NUMBER:  US 60/111,330
;  PRIOR FILING DATE:  1998-12-07
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  Microsoft Word
; SEQ ID NO 3
;    LENGTH: 11459
;    TYPE: DNA
;    ORGANISM: Venezuelan Equine Encephalitis Virus
;    FEATURE:
US-09-454-721A-3


  Query Match             48.7%;  Score 19;  DB 4;  Length 11459;
  Best Local Similarity   81.5%;  Pred. No. 47;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGC 28
           |||||  |||||||| || ||| ||||
Db    7795 GCCTCCCCCTTTCTGTTTTTGCGAGGC 7769


RESULT   3
US-09-181-336-12
; Sequence 12, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
;  APPLICANT: FREEMAN, Craig Geoffreey
;  APPLICANT: PARISH, Richard
;  APPLICANT: HAMDORF, Brenton James
;  APPLICANT: HULETT, Mark Darren
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
;  TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
;  FILE REFERENCE: 032505-004
;  CURRENT APPLICATION NUMBER: US/09/181,336A
;  CURRENT FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: AU PP0062
;  EARLIER FILING DATE: 1997-10-28
;  EARLIER APPLICATION NUMBER: AU PP0812
;  EARLIER FILING DATE: 1997-12-09
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;    LENGTH: 1713
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (46)..(1674)
;    FEATURE:
;    NAME/KEY: mat_peptide
;    LOCATION: (517)..(1674)
US-09-181-336-12


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1713;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     256 ttcctcatcctcctgggttctccaaagcttcgtacctt 293


RESULT   4
US-08-922-170B-9
; Sequence 9, Application US/08922170B
; Patent No. 5968822
;  GENERAL INFORMATION:
;    APPLICANT:  Iris Pecker, Israel Vlodavsky and Elena
;    APPLICANT:  Feinstein
;    TITLE OF INVENTION:  POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
;    TITLE OF INVENTION:  HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
;    TITLE OF INVENTION:  SAME IN TRANSDUCED CELLS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mark M. Friedman c/o Robert Sheinbein
;      STREET:  2940 Birchtree lane
;      CITY:  Silver Spring
;      STATE:  Maryland
;      COUNTRY:  United States of America
;      ZIP:  20906
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;      COMPUTER:  Twinhead* Slimnote-890TX
;      OPERATING SYSTEM:  MS DOS version 6.2,
;      OPERATING SYSTEM:  Windows version 3.11
;      SOFTWARE:  Word for Windows version 2.0 converted to
;      SOFTWARE:  an ASCI file
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/922,170B
;      FILING DATE:  2 SEP 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Friedmam, Mark M.
;      REGISTRATION NUMBER:  33,883
;      REFERENCE/DOCKET NUMBER:  910/1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  972-3-5625553
;      TELEFAX:  972-3-5625554
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1721
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-922-170B-9


  Query Match             48.2%;  Score 18.8;  DB 2;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310


RESULT   5
US-08-922-170B-11
; Sequence 11, Application US/08922170B
; Patent No. 5968822
;  GENERAL INFORMATION:
;    APPLICANT:  Iris Pecker, Israel Vlodavsky and Elena
```

```
;     APPLICANT:  Feinstein
;     TITLE OF INVENTION:  POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
;     TITLE OF INVENTION:  HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
;     TITLE OF INVENTION:  SAME IN TRANSDUCED CELLS
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Mark M. Friedman c/o Robert Sheinbein
;       STREET:  2940 Birchtree lane
;       CITY:  Silver Spring
;       STATE:  Maryland
;       COUNTRY:  United States of America
;       ZIP:  20906
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;       COMPUTER:  Twinhead* Slimnote-890TX
;       OPERATING SYSTEM:  MS DOS version 6.2,
;       OPERATING SYSTEM:  Windows version 3.11
;       SOFTWARE:  Word for Windows version 2.0 converted to
;       SOFTWARE:  an ASCI file
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/922,170B
;       FILING DATE:  2 SEP 1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Friedmam, Mark M.
;       REGISTRATION NUMBER:  33,883
;       REFERENCE/DOCKET NUMBER:  910/1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  972-3-5625553
;       TELEFAX:  972-3-5625554
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1721
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-922-170B-11


  Query Match             48.2%;  Score 18.8;  DB 2;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310


RESULT   6
US-09-071-739B-1
; Sequence 1, Application US/09071739B
; Patent No. 6177545
;  GENERAL INFORMATION:
;     APPLICANT:  Iris Pecker et al.
;     TITLE OF INVENTION:  HEPARANASE SPECIFIC MOLECULAR PROBES
;     TITLE OF INVENTION:  AND THEIR USE IN RESEARCH AND MEDICAL
;     TITLE OF INVENTION:  APPLICATIONS
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Mark M. Friedman c/o Anthony Castorina
;       STREET:  20001 Jefferson Davis Highway, Suite 207
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  United States of America
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;       COMPUTER:  Twinhead* Slimnote-890TX
;       OPERATING SYSTEM:  MS DOS version 6.2,
;       OPERATING SYSTEM:  Windows version 3.11
;       SOFTWARE:  Word for Windows version 2.0 converted to
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/071,739B
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/922,180
;       FILING DATE:  September 2, 1997
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Friedmam, Mark M.
;       REGISTRATION NUMBER:  33,883
;       REFERENCE/DOCKET NUMBER:  910/5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  972-3-5625553
;       TELEFAX:  972-3-5625554
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1721
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-071-739B-1


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310


RESULT   7
US-09-071-739B-3
; Sequence 3, Application US/09071739B
; Patent No. 6177545
;  GENERAL INFORMATION:
;     APPLICANT:  Iris Pecker et al.
;     TITLE OF INVENTION:  HEPARANASE SPECIFIC MOLECULAR PROBES
;     TITLE OF INVENTION:  AND THEIR USE IN RESEARCH AND MEDICAL
;     TITLE OF INVENTION:  APPLICATIONS
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Mark M. Friedman c/o Anthony Castorina
;       STREET:  20001 Jefferson Davis Highway, Suite 207
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  United States of America
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;       COMPUTER:  Twinhead* Slimnote-890TX
;       OPERATING SYSTEM:  MS DOS version 6.2,
;       OPERATING SYSTEM:  Windows version 3.11
;       SOFTWARE:  Word for Windows version 2.0 converted to
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/071,739B
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/922,180
;       FILING DATE:  September 2, 1997
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Friedmam, Mark M.
;       REGISTRATION NUMBER:  33,883
;       REFERENCE/DOCKET NUMBER:  910/5
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  972-3-5625553
;      TELEFAX:  972-3-5625554
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1721
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  linear
US-09-071-739B-3


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310


RESULT   8
US-09-260-038B-1
; Sequence 1, Application US/09260038B
; Patent No. 6348344
;   GENERAL INFORMATION:
;        APPLICANT: Maty Ayal-Hershkovitz et al.
;        TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
;                            EXPRESSING RECOMBINANT HEPARANASE
;                            AND METHODS OF PURIFYING SAME
;        NUMBER OF SEQUENCES: 25
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
;             STREET: 2001 Jefferson Davis Highway, Suite 207
;             CITY: Arlington
;             STATE: Virginia
;             COUNTRY: United States of America
;             ZIP: 22202
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
;             COMPUTER: Twinhead* Slimnote-890TX
;             OPERATING SYSTEM: MS DOS version 6.2,
;                               Windows version 3.11
;             SOFTWARE: Word for Windows version 2.0 converted to
;                       an ASCI file
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/260,038B
;             FILING DATE: 02-Mar-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/071,618
;             FILING DATE: May 1, 1998
;             APPLICATION NUMBER: 09/071,739
;             FILING DATE: May 1, 1998
;             APPLICATION NUMBER: 08/922,180
;             FILING DATE: September 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Friedmam, Mark M.
;             REGISTRATION NUMBER: 33,883
;             REFERENCE/DOCKET NUMBER: 910/16
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 972-3-5625553
;             TELEFAX: 972-3-5625554
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1721
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-260-038B-1


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310


RESULT   9
US-09-260-038B-3
; Sequence 3, Application US/09260038B
; Patent No. 6348344
;   GENERAL INFORMATION:
;        APPLICANT: Maty Ayal-Hershkovitz et al.
;        TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
;                            EXPRESSING RECOMBINANT HEPARANASE
;                            AND METHODS OF PURIFYING SAME
;        NUMBER OF SEQUENCES: 25
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
;             STREET: 2001 Jefferson Davis Highway, Suite 207
;             CITY: Arlington
;             STATE: Virginia
;             COUNTRY: United States of America
;             ZIP: 22202
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
;             COMPUTER: Twinhead* Slimnote-890TX
;             OPERATING SYSTEM: MS DOS version 6.2,
;                               Windows version 3.11
;             SOFTWARE: Word for Windows version 2.0 converted to
;                       an ASCI file
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/260,038B
;             FILING DATE: 02-Mar-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/071,618
;             FILING DATE: May 1, 1998
;             APPLICATION NUMBER: 09/071,739
;             FILING DATE: May 1, 1998
;             APPLICATION NUMBER: 08/922,180
;             FILING DATE: September 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Friedmam, Mark M.
;             REGISTRATION NUMBER: 33,883
;             REFERENCE/DOCKET NUMBER: 910/16
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 972-3-5625553
;             TELEFAX: 972-3-5625554
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1721
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-260-038B-3


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1721;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     273 TTCCTCATCCTCCTGGGTTCTCCAAAGCTTCGTACCTT 310
```

```
RESULT 10
US-09-181-336-14
; Sequence 14, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
;  APPLICANT: FREEMAN, Craig Geoffreey
;  APPLICANT: PARISH, Richard
;  APPLICANT: HAMDORF, Brenton James
;  APPLICANT: HULETT, Mark Darren
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
;  TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
;  FILE REFERENCE: 032505-004
;  CURRENT APPLICATION NUMBER: US/09/181,336A
;  CURRENT FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: AU PP0062
;  EARLIER FILING DATE: 1997-10-28
;  EARLIER APPLICATION NUMBER: AU PP0812
;  EARLIER FILING DATE: 1997-12-09
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;  LENGTH: 1723
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (52)..(1647)
US-09-181-336-14


  Query Match             48.2%;  Score 18.8;  DB 4;  Length 1723;
  Best Local Similarity   68.4%;  Pred. No. 35;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTT 38
           | |||| ||||    |||||| ||| ||  |  |||||
Db     262 ttcctcatcctcctgggttctccaaagcttcgtacctt 299


RESULT 11
US-08-019-870-2/c
; Sequence 2, Application US/08019870
; Patent No. 5336613
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  YOSHIMASA, SAITO
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  ISHII, YOSHINORI
;    TITLE OF INVENTION:  A NEW CEPHALOSPORIN C ACYLASE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/019,870
;      FILING DATE:  19930219
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5336613man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-791-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2322
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4..2322
US-08-019-870-2


  Query Match             47.7%;  Score 18.6;  DB 1;  Length 2325;
  Best Local Similarity   72.7%;  Pred. No. 47;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGA 34
           || ||| ||  || |||||||| | || || ||
Db    1959 GCTTCGCCCGGTCAGGTTCTGCGACGCGACCGA 1927


RESULT 12
US-08-019-870-4/c
; Sequence 4, Application US/08019870
; Patent No. 5336613
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  YOSHIMASA, SAITO
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  ISHII, YOSHINORI
;    TITLE OF INVENTION:  A NEW CEPHALOSPORIN C ACYLASE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/019,870
;      FILING DATE:  19930219
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5336613man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-791-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2322
```

```
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4..2322
US-08-019-870-4


  Query Match             47.7%;  Score 18.6;  DB 1;  Length 2325;
  Best Local Similarity   72.7%;  Pred. No. 47;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGA 34
           || ||| ||  || ||||||||  | || || ||
Db    1959 GCTTCGCCCGGTCAGGTTCTGCGACGCGACCGA 1927


RESULT  13
US-08-019-870-7/c
; Sequence 7, Application US/08019870
; Patent No. 5336613
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  YOSHIMASA, SAITO
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  ISHII, YOSHINORI
;    TITLE OF INVENTION:  A NEW CEPHALOSPORIN C ACYLASE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/019,870
;      FILING DATE:  19930219
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5336613man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-791-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2322
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4..2322
US-08-019-870-7


  Query Match             47.7%;  Score 18.6;  DB 1;  Length 2325;
  Best Local Similarity   72.7%;  Pred. No. 47;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGA 34
           || ||| ||  || ||||||||  | || || ||
Db    1959 GCTTCGCCCGGTCAGGTTCTGCGACGCGACCGA 1927


RESULT  14
US-08-019-870-10/c
; Sequence 10, Application US/08019870
; Patent No. 5336613
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  YOSHIMASA, SAITO
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  ISHII, YOSHINORI
;    TITLE OF INVENTION:  A NEW CEPHALOSPORIN C ACYLASE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/019,870
;      FILING DATE:  19930219
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5336613man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-791-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2322
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4..2322
US-08-019-870-10


  Query Match             47.7%;  Score 18.6;  DB 1;  Length 2325;
  Best Local Similarity   72.7%;  Pred. No. 47;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGA 34
           || ||| ||  || ||||||||  | || || ||
Db    1959 GCTTCGCCCGGTCAGGTTCTGCGACGCGACCGA 1927


RESULT  15
US-08-314-309A-20/c
; Sequence 20, Application US/08314309A
; Patent No. 5677141
;  GENERAL INFORMATION:
;    APPLICANT:  ISOGAI, TAKAO
;    APPLICANT:  FUKAGAWA, MASAO
;    APPLICANT:  IWAMI, MORITA
;    APPLICANT:  ARAMORI, ICHIRO
```

; APPLICANT: KOJO, HITOSHI
; TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM
; TITLE OF INVENTION: COMPOUND OR SALTS THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,309A
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,906
; FILING DATE: 21-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 567714lman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-863-0 CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2322
US-08-314-309A-20

Query Match 47.7%; Score 18.6; DB 1; Length 2325;
Best Local Similarity 72.7%; Pred. No. 47;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       2 GCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGA 34
           || ||| ||  || |||||||| | || || ||
Db    1959 GCTTCGCCCGGTCAGGTTCTGCGACGCGACCGA 1927
```

Search completed: June 6, 2002, 07:38:15
Job time: 5317 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 27.21 Seconds
(without alignments)
45.908 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	60.9	82	2	S35770	T-cell receptor al
2	38	59.4	356	2	AE2729	cytochrome oxidase
3	38	59.4	372	2	F97510	BA483f11.2.1 (cox1
4	38	59.4	803	1	IJCH3	integrin, band 3 p
5	37	57.8	223	2	T09926	cytidine deaminase
6	37	57.8	337	2	T09925	cytidine deaminase
7	36	56.2	173	1	NKVGCU	core protein p20 -
8	36	56.2	260	2	E84668	hypothetical prote
9	36	56.2	262	2	F87249	enoyl-CoA hydratas
10	36	56.2	315	1	S74441	iron(III) dicitrat
11	36	56.2	347	2	AB0961	cytochrome c-type
12	36	56.2	347	2	AE0787	cytochrome c-type
13	36	56.2	415	1	W2WLEP	E2 protein - Europ
14	36	56.2	606	2	E87318	hypothetical prote
15	35	54.7	152	2	S36550	E6 protein - human
16	35	54.7	254	2	JC5004	attacin A precurso
17	35	54.7	307	2	T09923	cytidine deaminase
18	35	54.7	336	2	A43260	glyceraldehyde-3-p
19	35	54.7	379	2	JN0013	synaptic vesicle m
20	35	54.7	741	2	AF2297	hypothetical prote
21	35	54.7	905	2	T40015	phosphatidylethano
22	35	54.7	2715	2	T13049	eyelid - fruit fly
23	34.5	53.9	2039	2	S64540	probable calcium c
24	34	53.1	148	2	S36532	E6 protein - human
25	34	53.1	183	2	E85182	hypothetical prote
26	34	53.1	192	2	A64835	probable membrane
27	34	53.1	192	2	F90757	hypothetical prote
28	34	53.1	192	2	D85621	hypothetical prote
29	34	53.1	199	2	S48362	probable membrane
30	34	53.1	208	2	AH2535	hypothetical prote
31	34	53.1	256	2	H69747	protein kinase hom
32	34	53.1	272	1	G69587	L-arabinose operon
33	34	53.1	301	2	T52549	probable thiosulfa
34	34	53.1	305	2	C85878	D-fructokinase [im
35	34	53.1	339	2	E83496	hypothetical prote
36	34	53.1	349	2	T06680	hypothetical prote
37	34	53.1	388	2	E65054	hypothetical prote
38	34	53.1	394	2	T41639	ion transporter -
39	34	53.1	412	2	T41141	probable caCA prot
40	34	53.1	419	2	C86223	hypothetical prote
41	34	53.1	422	1	A60503	sperm-binding glyc
42	34	53.1	423	2	S56672	probable 26S prote
43	34	53.1	423	2	B75503	sporulation protei
44	34	53.1	445	2	B49776	hypothetical prote
45	34	53.1	615	2	S77332	NADH dehydrogenase

ALIGNMENTS

RESULT 1
S35770
T-cell receptor alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 23-Jul-1999
C;Accession: S35770
R;Wedderburn, L.R.
submitted to the EMBL Data Library, June 1993
A;Reference number: S35769
A;Accession: S35770
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-82 <WED>
A;Cross-references: EMBL:Z22966; NID:g312155; PIDN:CAA80539.1; PID:g312156
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

Query Match 60.9%; Score 39; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GSARQLTF 13
||||||||
Db 60 GSARQLTF 67

RESULT 2
AE2729
cytochrome oxidase assembly factor cox15 [imported] - Agrobacterium tumefaciens (stra
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AE2729
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AE2729
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-356 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42251.1; PID:g17739647; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: cox15
A;Map position: circular chromosome

```
  Query Match             59.4%;  Score 38;  DB 2;  Length 356;
  Best Local Similarity   66.7%;  Pred. No. 16;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           |||| | | :||
Db      28 CLVLVGGATRLT 39


RESULT   3
F97510
BA483f11.2.1 (cox15 (yeast) homolog, cytochrome c oxidase assembly protein (isoform 1))
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: F97510
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: F97510
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87039.1; PID:g15156289; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_2290
A;Map position: circular chromosome

  Query Match             59.4%;  Score 38;  DB 2;  Length 372;
  Best Local Similarity   66.7%;  Pred. No. 17;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           |||| | | :||
Db      44 CLVLVGGATRLT 55


RESULT   4
IJCH3
integrin, band 3 precursor - chicken
N;Alternate names: CSAT antigen; JG22 antigen; RGD-receptor
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999
C;Accession: A23947
R;Tamkun, J.W.; DeSimone, D.W.; Fonda, D.; Patel, R.S.; Buck, C.; Horwitz, A.F.; Hynes,
Cell 46, 271-282, 1986
A;Title: Structure of integrin, a glycoprotein involved in the transmembrane linkage bet
A;Reference number: A23947; MUID:86245073
A;Accession: A23947
A;Molecule type: mRNA
A;Residues: 1-803 <TAM>
A;Cross-references: GB:M14049; NID:g212213; PIDN:AAA48926.1; PID:g212214
A;Experimental source: embryonic fibroblasts
A;Note: the amino end of the mature protein is blocked
C;Comment: Integrin, an integral plasma membrane complex of three (two in mammalian cell
C;Comment: This transmembrane complex may be the target of oncogenic transformation that
keleton. The cytoplasmic domain of this subunit contains a potential tyrosine-kinase pho
C;Comment: The extracellular domain of this protein, like many membrane receptors, conta
C;Superfamily: integrin beta chain; laminin-type EGF-like homology
C;Keywords: cell adhesion; cytoskeleton; duplication; extracellular matrix; glycoprotein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-803/Product: integrin, band 3 #status predicted <MAT>
F;25-733/Domain: extracellular #status predicted <EXT>
F;467-654/Region: cysteine-rich
F;482-564,565-650/Region: duplication
F;734-756/Domain: transmembrane #status predicted <MEM>
F;757-803/Domain: intracellular #status predicted <INT>
F;216,273,367,410,421,433,445,486,525,589,624,674/Binding site: carbohydrate (Asn) (cova
F;788/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             59.4%;  Score 38;  DB 1;  Length 803;
  Best Local Similarity   70.0%;  Pred. No. 37;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQ 10
           ||: ||||:|
Db      17 CLIWSGSAQQ 26


RESULT   5
T09926
cytidine deaminase (EC 3.5.4.5) CDA3 - Arabidopsis thaliana
N;Alternate names: protein T16L4.140
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C;Accession: T09926
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16897
A;Accession: T09926
A;Molecule type: DNA
A;Residues: 1-223 <BEV>
A;Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.140
A;Experimental source: cultivar Columbia; BAC clone T16L4
C;Genetics:
A;Gene: CDA3; ATSP:T16L4.140
A;Map position: 4
C;Superfamily: cdd protein
C;Keywords: hydrolase

  Query Match             57.8%;  Score 37;  DB 2;  Length 223;
  Best Local Similarity   63.6%;  Pred. No. 16;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||||| ::
Db     171 CLSLSGSAEEI 181


RESULT   6
T09925
cytidine deaminase (EC 3.5.4.5) CDA2 - Arabidopsis thaliana
N;Alternate names: protein T16L4.130
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Nov-2000
C;Accession: T09925; T52131
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16897
A;Accession: T09925
A;Molecule type: DNA
A;Residues: 1-337 <BEV>
A;Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.130
A;Experimental source: cultivar Columbia; BAC clone T16L4
R;Faivre-Nitschke, E.S.; Grienenberger, J.M.; Gualberto, J.M.
submitted to the EMBL Data Library, April 1998
A;Description: Cloning and characterisation of a cytidine deaminase gene family from
A;Reference number: Z25972
A;Accession: T52131
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-337 <FAI>
A;Cross-references: EMBL:AJ005811; PIDN:CAA06710.1
C;Genetics:
A;Gene: CDA2; ATSP:T16L4.130; cda2
A;Map position: 4
C;Superfamily: cdd protein
C;Keywords: hydrolase

  Query Match             57.8%;  Score 37;  DB 2;  Length 337;
  Best Local Similarity   63.6%;  Pred. No. 24;
```

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 CLVLSGSARQL 11
          || ||||| ::
Db    186 CLTLSGSAGEI 196
```

RESULT 7
NKVGCU
core protein p20 - cucumber necrosis virus
C;Species: cucumber necrosis virus
A;Note: host Cucumis sativus (cucumber)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Jul-1999
C;Accession: JA0133
R;Rochon, D.M.; Tremaine, J.H.
Virology 169, 251-259, 1989
A;Title: Complete nucleotide sequence of the cucumber necrosis virus genome.
A;Reference number: A94391; MUID:89204896
A;Accession: JA0133
A;Molecule type: genomic RNA
A;Residues: 1-173 <ROC>
A;Cross-references: GB:M25270; NID:g323338; PIDN:AAA42906.1; PID:g323343
C;Comment: The genome is a single-stranded, positive-sense RNA. It codes for 33K, 92K, 4
are core proteins.
C;Superfamily: tombusvirus core protein p19
C;Keywords: core protein

Query Match 56.2%; Score 36; DB 1; Length 173;
Best Local Similarity 63.6%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 CLVLSGSARQL 11
          || ||| :| |
Db    120 CLTLSGGSRTL 130
```

RESULT 8
E84668
hypothetical protein At2g27080 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: E84668
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84668
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <STO>
A;Cross-references: GB:AE002093; NID:g3885338; PIDN:AAC77866.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27080
A;Map position: 2

Query Match 56.2%; Score 36; DB 2; Length 260;
Best Local Similarity 81.8%; Pred. No. 29;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      2 LVLSGSARQLT 12
          ||||||  |||
Db    181 LVLSGSKIQLT 191
```

RESULT 9
F87249
enoyl-CoA hydratase/isomerase family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87249
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87249
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <STO>
A;Cross-references: GB:AE005673; NID:g13421092; PIDN:AAK21994.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0006
C;Superfamily: naphthoate synthase; enoyl-CoA hydratase homology

Query Match 56.2%; Score 36; DB 2; Length 262;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CLVLSGSAR 9
          |:||:|||:
Db     55 CIVLTGSAK 63
```

RESULT 10
S74441
iron(III) dicitrate transport system permease protein fecB - Synechocystis sp. (strai
N;Alternate names: protein slr1319
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: S74441
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74441
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-315 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16593.1; PID:g165
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: fecB
C;Superfamily: iron(III) dicitrate transport protein
C;Keywords: iron transport

Query Match 56.2%; Score 36; DB 1; Length 315;
Best Local Similarity 63.6%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 CLVLSGSARQL 11
          |||| | |:|:
Db     11 CLVLFGCAKQV 21
```

RESULT 11
AB0961
cytochrome c-type biogenesis protein H2 [imported] - Salmonella enterica subsp. enter
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AB0961
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; PMID:11677608
A;Accession: AB0961
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-347 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03182.1; PID:g16504815; GSPDB:GN00176
C;Genetics:
A;Gene: ccmH2

Query Match 56.2%; Score 36; DB 2; Length 347;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LVLSGSARQLT 12
||:||||| |
Db 14 LVISGSARATT 24

RESULT 12
AE0787
cytochrome c-type biogenesis protein H1 [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AE0787
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; PMID:11677608
A;Accession: AE0787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-347 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD07479.1; PID:g16503474; GSPDB:GN00176
C;Genetics:
A;Gene: ccmH1

Query Match 56.2%; Score 36; DB 2; Length 347;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LVLSGSARQLT 12
||:||||| |
Db 14 LVISGSARATT 24

RESULT 13
W2WLEP
E2 protein - European elk papillomavirus
C;Species: European elk papillomavirus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-May-2000
C;Accession: D29499; D94457; D94506
R;Ahola, H.; Bergman, P.; Stroem, A.C.; Moreno-Lopez, J.; Pettersson, U.
Gene 50, 195-205, 1986
A;Title: Organization and expression of the transforming region from the European elk pa
A;Reference number: A91567; MUID:87219878
A;Accession: D29499
A;Molecule type: DNA
A;Residues: 1-415 <AHO>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020
R;Eriksson, A.
unpublished results 1987, cited by GenBank
A;Reference number: A94457
A;Accession: D94457
A;Molecule type: DNA
A;Residues: 1-415 <ERI>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020
R;Pettersson, U.
submitted to GenBank, August 1987
A;Reference number: A94506
A;Accession: D94506
A;Molecule type: DNA
A;Residues: 1-415 <PET>
A;Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020
C;Superfamily: papillomavirus E2 protein
C;Keywords: early protein

Query Match 56.2%; Score 36; DB 1; Length 415;
Best Local Similarity 60.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQ 10
||:|||: |
Db 333 CLILSGNGNQ 342

RESULT 14
E87318
hypothetical protein CC0559 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87318
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87318
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-606 <STO>
A;Cross-references: GB:AE005673; NID:g13421751; PIDN:AAK22545.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0559

Query Match 56.2%; Score 36; DB 2; Length 606;
Best Local Similarity 58.3%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQLT 12
||| | |: ||
Db 501 CLVADGQAKALT 512

RESULT 15
S36550
E6 protein - human papillomavirus type 3
C;Species: human papillomavirus type 3
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C;Accession: S36550
R;Dellus, H.; Hofmann, B.
submitted to the EMBL Data Library, August 1993
A;Description: Primer-directed sequencing of human papillomavirus types.
A;Reference number: S36469
A;Accession: S36550
A;Molecule type: DNA
A;Residues: 1-152 <DEL>
A;Cross-references: EMBL:X74462; NID:g397005; PIDN:CAA52469.1; PID:g397006
C;Superfamily: papillomavirus E6 protein
C;Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match 54.7%; Score 35; DB 2; Length 152;
Best Local Similarity 46.2%; Pred. No. 27;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQLTF 13

```
          ||:: | ||:| :
Db      69 CLLVEGIARRLKY 81
```

Search completed: June 6, 2002, 06:16:22
Job time: 66 sec

05

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:16:26 ; Search time 13.46 Seconds
(without alignments)
37.396 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	59.4	803	1	ITB1_CHICK	P07228 gallus gall
2	36	56.2	173	1	VP19_CNV	P15184 cucumber ne
3	36	56.2	415	1	VE2_PAPVE	P11329 european el
4	35	54.7	152	1	VE6_HPV03	P36799 human papil
5	35	54.7	254	1	ATTA_TRINI	P50725 trichoplusi
6	35	54.7	336	1	G3P_CORGL	Q01651 corynebacte
7	35	54.7	379	1	VAT1_TORCA	P19333 torpedo cal
8	35	54.7	905	1	CHO2_SCHPO	O74787 schizosacch
9	34.5	53.9	2039	1	CCH1_YEAST	P50077 saccharomyc
10	34	53.1	148	1	VE6_HPV10	P36802 human papil
11	34	53.1	180	1	YCBW_ECOLI	P75862 escherichia
12	34	53.1	357	1	PMFE_PROMI	P53522 proteus mir
13	34	53.1	388	1	YGBK_ECOLI	Q46889 escherichia
14	34	53.1	419	1	PRSA_ARATH	O04019 arabidopsis
15	34	53.1	422	1	ZP3_MESAU	P23491 mesocricetu
16	34	53.1	423	1	PRSA_LYCES	P54776 lycopersico
17	34	53.1	424	1	PRSA_BRACM	O23894 brassica ca
18	34	53.1	429	1	PRSA_ORYSA	P46465 oryza sativ
19	34	53.1	442	1	FTSA_RHIME	O30994 rhizobium m
20	34	53.1	530	1	UDB8_RAT	Q62789 rattus norv
21	34	53.1	533	1	ST23_HUMAN	Q9upe1 homo sapien
22	34	53.1	838	1	HIS2_CANAL	O74712 candida alb
23	33.5	52.3	389	1	WN1B_HUMAN	O00744 homo sapien
24	33.5	52.3	389	1	WN1B_MOUSE	P48614 mus musculu
25	33	51.6	193	1	Y983_HAEIN	P43907 haemophilus
26	33	51.6	231	1	CASB_MOUSE	P10598 mus musculu
27	33	51.6	339	1	SRR_MOUSE	Q9qzx7 mus musculu
28	33	51.6	340	1	SRR_HUMAN	Q9gzt4 homo sapien
29	33	51.6	416	1	VE2_PAPVD	P03123 deer papill
30	33	51.6	443	1	FTSA_AGRTU	O30991 agrobacteri
31	33	51.6	494	1	G6PD_ACTAC	P77809 actinobacil
32	33	51.6	511	1	XASA_ECO57	P58229 escherichia
33	33	51.6	521	1	SR54_MYCLE	O33013 mycobacteri
34	33	51.6	530	1	UDBC_RAT	P36511 rattus norv
35	33	51.6	610	1	CHIT_STRPL	P11220 streptomyce
36	33	51.6	740	1	RNFC_ECO57	P58324 escherichia
37	33	51.6	740	1	RNFC_ECOLI	P77611 escherichia
38	33	51.6	1235	1	DPOL_PYRHO	O59610 pyrococcus
39	33	51.6	2715	1	G156_PARPR	P13837 paramecium
40	33	51.6	3396	1	POLG_DEN1S	P33478 d genome po
41	32.5	50.8	1500	1	CPSM_HUMAN	P31327 homo sapien
42	32.5	50.8	1500	1	CPSM_RAT	P07756 rattus norv
43	32	50.0	131	1	YRAN_ECOLI	P45465 escherichia
44	32	50.0	146	1	VE6_HPV28	P50802 human papil
45	32	50.0	188	1	RUVC_MYCTU	Q50627 mycobacteri

ALIGNMENTS

RESULT 1
ITB1_CHICK
ID ITB1_CHICK STANDARD; PRT; 803 AA.
AC P07228;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin beta-1 precursor (CSAT antigen) (JG22 antigen) (RGD-
DE receptor).
GN ITGB1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic fibroblast;
RX MEDLINE=86245073; PubMed=3487386;
RA Tamkun J.W., Desimone D.W., Fonda D., Patel R.S., Buck C.,
RA Horwitz A.F., Hynes R.O.;
RT "Structure of integrin, a glycoprotein involved in the transmembrane
RT linkage between fibronectin and actin.";
RL Cell 46:271-282(1986).
CC -!- FUNCTION: INTEGRINS ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-10/BETA-
CC 1 AND ALPHA-11/BETA-1 ARE RECEPTORS FOR COLLAGEN. INTEGRINS ALPHA-
CC 1/BETA-1 AND ALPHA-2/BETA-2 RECOGNIZE THE PROLINE-HYDROXYLATED
CC SEQUENCE G-F-P-G-E-R IN COLLAGEN. INTEGRINS ALPHA-2/BETA-1, ALPHA-
CC 3/BETA-1, ALPHA-4/BETA-1, ALPHA-5/BETA-1, ALPHA-8/BETA-1, ALPHA-
CC 10/BETA-1, ALPHA-11/BETA-1 AND ALPHA-V/BETA-1 ARE RECEPTORS FOR
CC FIBRONECTIN. ALPHA-4/BETA-1 RECOGNIZES ONE OR MORE DOMAINS WITHIN
CC THE ALTERNATIVELY SPLICED CS-1 AND CS-5 REGIONS OF FIBRONECTIN.
CC INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRINOGEN. INTEGRIN
CC ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-6/BETA-1 AND ALPHA-7/BETA-1
CC ARE RECEPTORS FOR LAMIMIN. INTEGRIN ALPHA-4/BETA-1 IS A RECEPTOR
CC FOR VCAM1 AND RECOGNIZES THE SEQUENCE Q-I-D-S IN VCAM1. INTEGRIN
CC ALPHA-9/BETA-1 IS A RECEPTOR FOR VCAM1, CYTOTACTIN AND
CC OSTEOPONTIN. IT RECOGNIZES THE SEQUENCE A-E-I-D-G-I-E-L IN
CC CYTOTACTIN. INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR EPILIGRIN
CC AND THROMBOSPONDIN. INTEGRIN ALPHA-V/BETA-1 IS A RECEPTOR FOR
CC VITRONECTIN. BETA-1 INTEGRINS RECOGNIZE THE SEQUENCE R-G-D IN A
CC WIDE ARRAY OF LIGANDS.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-1
CC ASSOCIATES WITH EITHER ALPHA-1, ALPHA-2, ALPHA-3, ALPHA-4, ALPHA-
CC 5, ALPHA-6, ALPHA-7, ALPHA-8, ALPHA-9, ALPHA-10, ALPHA-11 OR
CC ALPHA-V.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC ---

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14049; AAA48926.1; -.
DR   PIR; A23947; IJCH3.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002369; Integrin_B.
DR   InterPro; IPR001169; Integrin_beta_C.
DR   InterPro; IPR003659; PSI.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00362; integrin_B; 1.
DR   PRINTS; PR01186; INTEGRINB.
DR   ProDom; PD001811; Integrin_B; 1.
DR   SMART; SM00187; INB; 1.
DR   SMART; SM00423; PSI; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00243; INTEGRIN_BETA; 3.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
KW   Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW   Repeat; Phosphorylation; Signal.
FT   SIGNAL       1    24       BY SIMILARITY.
FT   CHAIN       25   803       INTEGRIN BETA-1.
FT   DOMAIN      25   733       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   734   756       POTENTIAL.
FT   DOMAIN     757   803       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     144   382       VWFA-LIKE.
FT   DOMAIN     471   640       4 CYSTEINE-RICH TANDEM REPEATS.
FT   REPEAT     471   520       I.
FT   REPEAT     521   564       II.
FT   REPEAT     565   603       III.
FT   REPEAT     604   640       IV.
FT   DISULFID    31   469       BY SIMILARITY.
FT   DISULFID    39    49       BY SIMILARITY.
FT   DISULFID    42    79       BY SIMILARITY.
FT   DISULFID    52    68       BY SIMILARITY.
FT   DISULFID   211   217       BY SIMILARITY.
FT   DISULFID   265   305       BY SIMILARITY.
FT   DISULFID   405   419       BY SIMILARITY.
FT   DISULFID   439   696       BY SIMILARITY.
FT   DISULFID   467   471       BY SIMILARITY.
FT   DISULFID   482   494       BY SIMILARITY.
FT   DISULFID   491   530       BY SIMILARITY.
FT   DISULFID   496   505       BY SIMILARITY.
FT   DISULFID   507   521       BY SIMILARITY.
FT   DISULFID   536   541       BY SIMILARITY.
FT   DISULFID   538   573       BY SIMILARITY.
FT   DISULFID   543   558       BY SIMILARITY.
FT   DISULFID   560   565       BY SIMILARITY.
FT   DISULFID   579   584       BY SIMILARITY.
FT   DISULFID   581   612       BY SIMILARITY.
FT   DISULFID   586   595       BY SIMILARITY.
FT   DISULFID   597   604       BY SIMILARITY.
FT   DISULFID   618   623       BY SIMILARITY.
FT   DISULFID   620   666       BY SIMILARITY.
FT   DISULFID   625   635       BY SIMILARITY.
FT   DISULFID   638   641       BY SIMILARITY.
FT   DISULFID   645   654       BY SIMILARITY.
FT   DISULFID   651   728       BY SIMILARITY.
FT   DISULFID   670   704       BY SIMILARITY.
FT   MOD_RES     25    25       BLOCKED.
FT   MOD_RES    788   788       PHOSPHORYLATION (BY TYR-KINASES)
FT                               (POTENTIAL).
FT   CARBOHYD   216   216       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   273   273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   367   367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   410   410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   421   421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   433   433       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   445   445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   486   486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   525   525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   589   589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   624   624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   674   674       N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ   SEQUENCE   803 AA;  88553 MW;  2F6FEFCDF2C80457 CRC64;

  Query Match             59.4%;  Score 38;  DB 1;  Length 803;
  Best Local Similarity   70.0%;  Pred. No. 14;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQ 10
           ||: ||||:|
Db      17 CLIWSGSAQQ 26


RESULT   2
VP19_CNV
ID   VP19_CNV       STANDARD;      PRT;   173 AA.
AC   P15184;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   Core protein P19 (P20).
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204896; PubMed=2705296;
RA   Rochon D.M., Tremaine J.H.;
RT   "Complete nucleotide sequence of the cucumber necrosis virus genome.";
RL   Virology 169:251-259(1989).
CC   -!- SIMILARITY: TO OTHER TOMBUSVIRUSES CORE PROTEIN P19.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25270; AAA42906.1; -.
DR   PIR; JA0133; NKVGCU.
KW   Core protein.
SQ   SEQUENCE   173 AA;  19808 MW;  D05F870D4AA4C80F CRC64;

  Query Match             56.2%;  Score 36;  DB 1;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 7;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130


RESULT   3
VE2_PAPVE
ID   VE2_PAPVE      STANDARD;      PRT;   415 AA.
AC   P11329;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Probable regulatory protein E2.
GN   E2.
OS   European elk papillomavirus (EEPV).
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87219878; PubMed=3034730;
RA   Ahola H., Bergman P., Stroem A.C., Moreno-Lopez J., Petterson U.;
RT   "Organization and expression of the transforming region from the
```

```
RT   European elk papillomavirus (EEPV).";
RL   Gene 50:195-205(1986).
CC   -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC       IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
CC       IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC       ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC       WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC       BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC       INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC       REPLICATION.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15953; AAA66854.1; -.
DR   PIR; D29499; W2WLEP.
DR   HSSP; P03122; 2BOP.
DR   InterPro; IPR001866; E2_N.
DR   InterPro; IPR000427; Early2_C.
DR   Pfam; PF00511; E2_C; 1.
DR   Pfam; PF00508; E2_N; 1.
DR   ProDom; PD000672; Early2_C; 1.
DR   ProDom; PD000678; E2_N; 1.
KW   Early protein; Transcription regulation; Activator; DNA-binding;
KW   Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ   SEQUENCE   415 AA;  46245 MW;  0B7BF6AEEF0AC80C CRC64;


  Query Match             56.2%;  Score 36;  DB 1;  Length 415;
  Best Local Similarity   60.0%;  Pred. No. 17;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQ 10
          ||:|||:  |
Db    333 CLILSGNGNQ 342


RESULT   4
VE6_HPV03
ID   VE6_HPV03      STANDARD;      PRT;   152 AA.
AC   P36799; Q81960;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   E6 protein.
GN   E6.
OS   Human papillomavirus type 3.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10614;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94265501; PubMed=8205838;
RA   Delius H., Hofmann B.;
RT   "Primer-directed sequencing of human papillomavirus types.";
RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   -!- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE
CC       STRANDED DNA (IN VITRO).
CC   -!- SUBCELLULAR LOCATION: Nuclear matrix-associated.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74462; CAA52469.1; -.
DR   EMBL; X74462; CAA52470.1; ALT_INIT.
DR   PIR; S36550; S36550.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
KW   Early protein; DNA-binding; Nuclear protein; Zinc-finger.
FT   ZN_FING      33     69       POTENTIAL.
FT   ZN_FING     106    142       POTENTIAL.
SQ   SEQUENCE   152 AA;  17926 MW;  50E869412755862F CRC64;


  Query Match             54.7%;  Score 35;  DB 1;  Length 152;
  Best Local Similarity   46.2%;  Pred. No. 9.6;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          ||:: | ||:| :
Db     69 CLLVEGIARRLKY 81


RESULT   5
ATTA_TRINI
ID   ATTA_TRINI     STANDARD;      PRT;   254 AA.
AC   P50725;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Attacin A precursor.
OS   Trichoplusia ni (Cabbage looper).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Plusiinae; Trichoplusia.
OX   NCBI_TaxID=7111;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97045820; PubMed=8890742;
RA   Kang D., Lundstroem A., Steiner H.;
RT   "Trichoplusia ni attacin A, a differentially displayed insect gene
RT   coding for an antibacterial protein.";
RL   Gene 174:245-249(1996).
CC   -!- FUNCTION: ATTACINS ARE HEMOLYMPH ANTIBACTERIAL PROTEINS.
CC   -!- SIMILARITY: BELONGS TO THE ATTACIN/SARCOTOXIN II FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U46130; AAC47327.1; -.
KW   Insect immunity; Antibiotic; Hemolymph; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19     62       POTENTIAL.
FT   CHAIN        63    254       ATTACIN A.
SQ   SEQUENCE   254 AA;  27632 MW;  02B24A24AE4AC82A CRC64;


  Query Match             54.7%;  Score 35;  DB 1;  Length 254;
  Best Local Similarity   66.7%;  Pred. No. 16;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 LVLSGSARQLTF 13
          ||:| ||| | |
Db     12 LVVSASARYLVF 23


RESULT   6
G3P_CORGL
```

```
ID   G3P_CORGL      STANDARD;      PRT;   336 AA.
AC   Q01651;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
GN   GAP.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC   Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13059 / AS019;
RX   MEDLINE=93015645; PubMed=1400158;
RA   Eikmanns B.J.;
RT   "Identification, sequence analysis, and expression of a
RT   Corynebacterium glutamicum gene cluster encoding the three glycolytic
RT   enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate
RT   kinase, and triosephosphate isomerase.";
RL   J. Bacteriol. 174:6076-6086(1992).
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC       NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC   -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X59403; CAA42045.1; -.
DR   PIR; S23910; S23910.
DR   PIR; A43260; A43260.
DR   HSSP; P00362; 1GD1.
DR   InterPro; IPR000173; GAP_DH.
DR   Pfam; PF00044; gpdh; 1.
DR   Pfam; PF02800; gpdh_C; 1.
DR   PRINTS; PR00078; G3PDHDRGNASE.
DR   PROSITE; PS00071; GAPDH; 1.
KW   Glycolysis; Oxidoreductase; NAD.
FT   BINDING     153    153       GLYCERALDEHYDE 3-PHOSPHATE.
FT   ACT_SITE    180    180       ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ   SEQUENCE   336 AA;  36199 MW;  5514A0A0CF078219 CRC64;
```

```
  Query Match             54.7%;  Score 35;  DB 1;  Length 336;
  Best Local Similarity   63.6%;  Pred. No. 22;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 VLSGSARQLTF 13
           |::|||  |||
Db     237 VITGSATDLTF 247
```

```
RESULT   7
VAT1_TORCA
ID   VAT1_TORCA     STANDARD;      PRT;   379 AA.
AC   P19333;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Synaptic vesicle membrane protein VAT-1.
OS   Torpedo californica (Pacific electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX   NCBI_TaxID=7787;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Electric lobe;
RX   MEDLINE=90166593; PubMed=2483112;
RA   Linial M., Miller K., Scheller R.H.;
RT   "VAT-1: an abundant membrane protein from Torpedo cholinergic
RT   synaptic vesicles.";
RL   Neuron 2:1265-1273(1989).
CC   -!- FUNCTION: MAY PLAY A CENTRAL ROLE IN THE FUNCTIONS MEDIATED BY
CC       SPECIFIC CLASSES OF SYNAPTIC VESICLES.
CC   -!- SUBCELLULAR LOCATION: Membrane.
CC   -!- TISSUE SPECIFICITY: CHOLINERGIC SYNAPTIC VESICLES.
CC   -!- MISCELLANEOUS: SYNTHESIZED IN THE NEURONAL CELL BODIES AND
CC       TRANSPORTED TO THE TERMINALS. EACH VESICLE CONTAINS APPROXIMATELY
CC       28 MOLECULES OF VAT-1.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY. STRONG, TO MAMMALIAN
CC       VAT-1 HOMOLOGS.
DR   PIR; JN0013; JN0013.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW   Oxidoreductase; Zinc; Synapse; Membrane; Phosphorylation.
FT   MOD_RES     273    273       PHOSPHORYLATION (POTENTIAL).
SQ   SEQUENCE   379 AA;  41593 MW;  FA4ADA17E657F09C CRC64;
```

```
  Query Match             54.7%;  Score 35;  DB 1;  Length 379;
  Best Local Similarity   72.7%;  Pred. No. 25;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLT 12
           ||| ||| |:|
Db     257 LVLYGSANQVT 267
```

```
RESULT   8
CHO2_SCHPO
ID   CHO2_SCHPO     STANDARD;      PRT;   905 AA.
AC   O74787; P87301;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17) (PEAMT).
GN   CHO2 OR SPBC26H8.03.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 130-905 FROM N.A.
RA   Kanipes M.I., Henry S.A.;
RT   "The Schizosaccharomyces pombe cho2+ gene encodes a
RT   phosphatidylethanolamine methyltransferase.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine +
CC       phosphatidylethanolamine = S-adenosyl-L-homocysteine +
CC       phosphatidyl-N-methylethanolamine.
CC   -!- PATHWAY: FIRST STEP IN PHOSPHATIDYLETHANOLAMINE METHYLATION
CC       PATHWAY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL031743; CAA21095.1; -.
DR   EMBL; AF004113; AAB61410.1; -.
KW   Phospholipid biosynthesis; Transferase; Methyltransferase.
FT   CONFLICT    133    133       C -> G (IN REF. 2).
FT   CONFLICT    137    137       W -> G (IN REF. 2).
FT   CONFLICT    149    149       L -> Q (IN REF. 2).
SQ   SEQUENCE   905 AA;  102761 MW;  E46CF4AEC0CFE889 CRC64;


  Query Match             54.7%;  Score 35;  DB 1;  Length 905;
  Best Local Similarity   58.3%;  Pred. No. 60;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           | :  ||||:||
Db     846 CDISEGSARKLT 857


RESULT   9
CCH1_YEAST
ID   CCH1_YEAST     STANDARD;      PRT;  2039 AA.
AC   P50077;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Probable calcium-channel protein.
GN   CCH1 OR YGR217W OR G8501.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE OF 1-1360 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=97435481; PubMed=9290212;
RA   Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT   "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT   chromosome VII.";
RL   Yeast 13:1077-1090(1997).
RN   [2]
RP   SEQUENCE OF 1184-2039 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=96267763; PubMed=8701610;
RA   van der Aart Q.J.M., Kleine K., Steensma H.Y.;
RT   "Sequence analysis of the 43 kb CRM1-YLM9-PET54-DIE2-SMI1-PHO81-YHB4-
RT   PFK1 region from the right arm of Saccharomyces cerevisiae chromosome
RT   VII.";
RL   Yeast 12:385-390(1996).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98001560; PubMed=9343395;
RA   Paidhungat M., Garrett S.;
RT   "A homolog of mammalian, voltage-gated calcium channels mediates
RT   yeast pheromone-stimulated Ca2+ uptake and exacerbates the cdc1(Ts)
RT   growth defect.";
RL   Mol. Cell. Biol. 17:6339-6347(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: SOME, TO MAMMALIAN SODIUM AND CALCIUM CHANNELS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z73002; CAA97244.1; -.
DR   EMBL; Z73003; CAA97245.1; -.
DR   EMBL; X87941; CAA61165.1; -.
DR   SGD; S0003449; CCH1.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel.
FT   TRANSMEM    346    366       POTENTIAL.
FT   TRANSMEM    384    404       POTENTIAL.
FT   TRANSMEM    563    583       POTENTIAL.
FT   TRANSMEM    658    678       POTENTIAL.
FT   TRANSMEM    691    711       POTENTIAL.
FT   TRANSMEM    766    786       POTENTIAL.
FT   TRANSMEM    809    829       POTENTIAL.
FT   TRANSMEM    841    861       POTENTIAL.
FT   TRANSMEM    904    924       POTENTIAL.
FT   TRANSMEM    942    962       POTENTIAL.
FT   TRANSMEM    978    998       POTENTIAL.
FT   TRANSMEM   1207   1227       POTENTIAL.
FT   TRANSMEM   1247   1267       POTENTIAL.
FT   TRANSMEM   1277   1297       POTENTIAL.
FT   TRANSMEM   1340   1360       POTENTIAL.
FT   TRANSMEM   1408   1428       POTENTIAL.
FT   TRANSMEM   1452   1472       POTENTIAL.
FT   TRANSMEM   1529   1549       POTENTIAL.
FT   TRANSMEM   1554   1574       POTENTIAL.
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   TRANSMEM   1618   1638       POTENTIAL.
FT   TRANSMEM   1654   1674       POTENTIAL.
FT   TRANSMEM   1748   1768       POTENTIAL.
FT   CONFLICT   1185   1185       R -> Q (IN REF. 2).
FT   CONFLICT   1203   1203       I -> N (IN REF. 2).
SQ   SEQUENCE   2039 AA;  234597 MW;  800B3825D6C6E527 CRC64;


  Query Match             53.9%;  Score 34.5;  DB 1;  Length 2039;
  Best Local Similarity   61.5%;  Pred. No. 1.7e+02;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSGSARQLTF 13
           || :| :||| ||
Db    1318 CLTISNTARQ-TF 1329


RESULT   10
VE6_HPV10
ID   VE6_HPV10      STANDARD;      PRT;   148 AA.
AC   P36802;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   E6 protein.
GN   E6.
OS   Human papillomavirus type 10.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10603;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94265501; PubMed=8205838;
RA   Delius H., Hofmann B.;
RT   "Primer-directed sequencing of human papillomavirus types.";
RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   -!- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE
CC       STRANDED DNA (IN VITRO).
CC   -!- SUBCELLULAR LOCATION: Nuclear matrix-associated.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74465; CAA52489.1; -.
DR   PIR; S36532; S36532.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
KW   Early protein; DNA-binding; Nuclear protein; Zinc-finger.
FT   ZN_FING      29     65       POTENTIAL.
FT   ZN_FING     102    138       POTENTIAL.
SQ   SEQUENCE   148 AA;  17563 MW;  EFCA68C51E61DB1A CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 148;
  Best Local Similarity   46.2%;  Pred. No. 15;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          ||:| |  |:| :
Db     65 CLLLQGIVRRLKY 77


RESULT  11
YCBW_ECOLI
ID   YCBW_ECOLI     STANDARD;      PRT;   180 AA.
AC   P75862;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ycbW.
GN   YCBW OR B0946.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000196; AAC74032.1; ALT_INIT.
DR   EMBL; D90732; BAA35701.1; ALT_INIT.
DR   EcoGene; EG13715; ycbW.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   180 AA;  20589 MW;  7A8CC66ECE2D1B87 CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 180;
  Best Local Similarity   54.5%;  Pred. No. 18;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQL 11
          |:|::| | ||
Db    144 CVVIAGRAMQL 154


RESULT  12
PMFE_PROMI
ID   PMFE_PROMI     STANDARD;      PRT;   357 AA.
AC   P53522;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Putative minor fimbrial subunit pmfE precursor.
GN   PMFE.
OS   Proteus mirabilis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=584;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HI4320;
RX   MEDLINE=95047519; PubMed=7959033;
RA   Massad G., Mobley H.L.T.;
RT   "Genetic organization and complete sequence of the Proteus mirabilis
RT   pmf fimbrial operon.";
RL   Gene 150:101-104(1994).
CC   -!- SUBCELLULAR LOCATION: Fimbria.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35428; CAA84592.1; -.
KW   Fimbria; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    357       PUTATIVE MINOR FIMBRIAL SUBUNIT PMFE.
SQ   SEQUENCE   357 AA;  38876 MW;  59AD7E566D4899AA CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 357;
  Best Local Similarity   63.6%;  Pred. No. 36;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 VLSGSARQLTF 13
          : |||: ||||
Db    154 ISSGSSGQLTF 164


RESULT  13
YGBK_ECOLI
ID   YGBK_ECOLI     STANDARD;      PRT;   388 AA.
AC   Q46889;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ygbK.
GN   YGBK OR B2737.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI1011.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29579; AAA69247.1; -.
DR   EMBL; AE000357; AAC75779.1; -.
DR   EcoGene; EG13105; ygbK.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   388 AA;  41339 MW;  5824A120E9393892 CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 388;
  Best Local Similarity   63.6%;  Pred. No. 40;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLT 12
           :|||||  |:|
Db     253 VVLSGSCSQMT 263


RESULT  14
PRSA_ARATH
ID   PRSA_ARATH     STANDARD;      PRT;   419 AA.
AC   O04019;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   26S protease regulatory subunit 6A homolog (TAT-binding protein
DE   homolog 1) (TBP-1).
GN   AT1G09100 OR F7G19.2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC000106; AAB70397.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003960; AAA_sub.
DR   InterPro; IPR003959; AAA_subfam.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND     207    214       ATP (POTENTIAL).
SQ   SEQUENCE   419 AA;  46663 MW;  BD6213676FBCF44A CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 419;
  Best Local Similarity   53.8%;  Pred. No. 43;
  Matches    7;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLTF 13
           |:||  | ||  |
Db     101 CVVLKTSTRQTIF 113


RESULT  15
ZP3_MESAU
ID   ZP3_MESAU      STANDARD;      PRT;   422 AA.
AC   P23491;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C).
GN   ZP3.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=91078540; PubMed=2257975;
RA   Kinloch R.A., Ruiz-Seller B., Wassarman P.M.;
RT   "Genomic organization and polypeptide primary structure of zona
RT   pellucida glycoprotein hZP3, the hamster sperm receptor.";
RL   Dev. Biol. 142:414-421(1990).
CC   -!- FUNCTION: FUNCTIONS AS A SPERM-RECEPTOR. IT IS RESPONSIBLE FOR
CC       SPERM-ADHESION TO THE ZONA PELLUCIDA, AND MAY CONTRIBUTE TO THE
CC       SPECIES-SPECIFICITY OF THE INSEMINATION.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- TISSUE SPECIFICITY: OOCYTES.
CC   -!- DEVELOPMENTAL STAGE: GROWING OOCYTES.
CC   -!- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 1 ZP DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63629; AAA37079.1; -.
DR   InterPro; IPR001507; zona_pellucida.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    422       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    386       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    387    407       POTENTIAL.
FT   DOMAIN      408    422       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    306       ZP.
FT   DOMAIN      119    158       PRO-RICH.
FT   DOMAIN      208    257       PRO-RICH.
FT   CARBOHYD    146    146       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   422 AA;  45827 MW;  D0F95BE7FF8E7E01 CRC64;


  Query Match             53.1%;  Score 34;  DB 1;  Length 422;
  Best Local Similarity   60.0%;  Pred. No. 43;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQ 10
             ||:| | |:|
Db      10 CLLLCGGAKQ 19


Search completed: June  6, 2002, 06:22:27
Job time: 361 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:51 ; Search time 41.63 Seconds
(without alignments)
54.022 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	59.4	222	2	Q9F598	Q9f598 agrobacteri
2	37	57.8	223	10	Q9S847	Q9s847 arabidopsis
3	37	57.8	322	10	Q9ZT33	Q9zt33 arabidopsis
4	37	57.8	337	10	O65571	O65571 arabidopsis
5	37	57.8	337	10	Q9SWZ4	Q9swz4 arabidopsis
6	37	57.8	552	12	Q91SH2	Q91sh2 guinea pig
7	36	56.2	173	12	Q9IES5	Q9ies5 cucumber ne
8	36	56.2	173	12	Q9IES9	Q9ies9 cucumber ne
9	36	56.2	173	12	Q9IES8	Q9ies8 cucumber ne
10	36	56.2	173	12	Q9IES7	Q9ies7 cucumber ne
11	36	56.2	173	12	Q9IES6	Q9ies6 cucumber ne
12	36	56.2	243	10	Q9ATM2	Q9atm2 zea mays (m
13	36	56.2	260	10	Q9ZVD2	Q9zvd2 arabidopsis
14	36	56.2	262	16	Q9AC55	Q9ac55 caulobacter
15	36	56.2	315	16	P72593	P72593 synechocyst
16	36	56.2	606	16	Q9AAN8	Q9aan8 caulobacter
17	35	54.7	199	16	Q98AR4	Q98ar4 rhizobium l
18	35	54.7	257	2	Q9AHX8	Q9ahx8 pseudomonas
19	35	54.7	292	12	Q91LC7	Q91lc7 white spot
20	35	54.7	292	12	Q918H9	Q918h9 white spot
21	35	54.7	307	10	Q9SU87	Q9su87 arabidopsis
22	35	54.7	308	10	Q9SYU5	Q9syu5 arabidopsis
23	35	54.7	316	12	Q83110	Q83110 mouse adeno
24	35	54.7	320	5	O96381	O96381 hyphantria
25	35	54.7	329	10	Q9ZT35	Q9zt35 arabidopsis
26	35	54.7	357	3	Q9HGK3	Q9hgk3 debaryomyce
27	35	54.7	364	5	Q95SZ9	Q95sz9 drosophila
28	35	54.7	386	5	P91632	P91632 drosophila
29	35	54.7	436	2	Q9F839	Q9f839 micromonosp
30	35	54.7	1543	12	Q91ND9	Q91nd9 cactus viru
31	35	54.7	2703	5	Q9VEG7	Q9veg7 drosophila
32	35	54.7	2715	5	O61603	O61603 drosophila
33	35	54.7	3187	5	Q9BLV5	Q9blv5 leishmania
34	34	53.1	147	5	Q962D5	Q962d5 drosophila
35	34	53.1	183	10	Q9SUL0	Q9sul0 arabidopsis
36	34	53.1	199	3	Q03885	Q03885 saccharomyc
37	34	53.1	217	2	P70799	P70799 azotobacter
38	34	53.1	227	3	Q02489	Q02489 saccharomyc
39	34	53.1	248	2	Q9KJB0	Q9kjb0 flavobacter
40	34	53.1	248	2	Q9KJA8	Q9kja8 flavobacter
41	34	53.1	248	2	Q9KJA7	Q9kja7 flavobacter
42	34	53.1	256	16	O31435	O31435 bacillus su
43	34	53.1	272	16	P94526	P94526 bacillus su
44	34	53.1	281	2	Q93S33	Q93s33 prevotella
45	34	53.1	301	2	O50573	O50573 bacillus ps

ALIGNMENTS

RESULT 1
Q9F598
ID Q9F598 PRELIMINARY; PRT; 222 AA.
AC Q9F598;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE RIORF157 PROTEIN.
GN RIORF157.
OS Agrobacterium rhizogenes.
OG Plasmid pRi1724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRi1724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1):Construction of linking library
RT and physical map of pRi1724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;

```
RX   MEDLINE=20241294; PubMed=10780382;
RA   Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA   Yoshida K.;
RT   "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
RT   its flanking regions of pRi1724 in Japanese Agrobacterium
RT   rhizogenes.";
RL   Nucleic Acids Symp. Ser. 42:67-68(1999).
DR   EMBL; AP002086; BAB16276.1; -.
DR   InterPro; IPR000560; His_acid_phosphtse.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW   Plasmid.
SQ   SEQUENCE   222 AA;  23349 MW;  255F18D47EB07AD2 CRC64;


  Query Match             59.4%;  Score 38;  DB 2;  Length 222;
  Best Local Similarity   77.8%;  Pred. No. 25;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSAR 9
            ||||||:|:
Db      14 CLVLSGTAK 22


RESULT   2
Q9S847
ID   Q9S847      PRELIMINARY;      PRT;   223 AA.
AC   Q9S847;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYTIDINE DEAMINASE 3 (CDA3) (EC 3.5.4.5).
GN   T16L4.140 OR DESF OR AT4G29630.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Rose M., Hempel S., Entian K.-D., Bancroft I., Mewes H.W.,
RA   Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Sanchez H., Schuster W.;
RT   "Cytidine deaminases in Arabidopsis thaliana: a gene family of eight
RT   members are located within a 24 kb region.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL079344; CAB45323.1; -.
DR   EMBL; AF121877; AAD30446.1; -.
DR   EMBL; AL161575; CAB79721.1; -.
DR   HSSP; P13652; 1ALN.
DR   InterPro; IPR002125; dCMP_cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
DR   PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW   Hydrolase.
SQ   SEQUENCE   223 AA;  24178 MW;  028CAD0778A50362 CRC64;


  Query Match             57.8%;  Score 37;  DB 10;  Length 223;
  Best Local Similarity   63.6%;  Pred. No. 39;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
            || ||||| ::
Db     171 CLSLSGSAEEI 181


RESULT   3
Q9ZT33
ID   Q9ZT33      PRELIMINARY;      PRT;   322 AA.
AC   Q9ZT33;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYTIDINE DEAMINASE 3 (EC 3.5.4.5).
GN   CDA3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Faivre-Nitschke S.E., Grienenberger J.M., Gualberto J.M.;
RT   "Cloning and characterisation of a cytidine deaminase gene family from
RT   Arabidopsis thaliana.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF080676; AAC69568.1; -.
DR   HSSP; P13652; 1ALN.
DR   InterPro; IPR002125; dCMP_cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
DR   PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW   Hydrolase.
SQ   SEQUENCE   322 AA;  35183 MW;  4985CAB806E3AE35 CRC64;


  Query Match             57.8%;  Score 37;  DB 10;  Length 322;
  Best Local Similarity   63.6%;  Pred. No. 58;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
            || ||||| ::
Db     171 CLSLSGSAEEI 181


RESULT   4
O65571
ID   O65571      PRELIMINARY;      PRT;   337 AA.
AC   O65571;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CYTIDINE DEAMINASE (EC 3.5.4.5) (CYTIDINE DEAMINASE 2) (CDA2).
GN   CDA2 OR DESE OR T16L4.130 OR AT4G29620.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Faivre Nitschke E.S., Grienenberger J.M., Gualberto J.M.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Sanchez H., Schuster W.;
RT   "Cytidine deaminases in Arabidopsis thaliana: a gene family of eight
RT   members are located within a 24 kb region.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Rose M., Hempel S., Entian K.-D., Bancroft I., Mewes H.W.,
RA   Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ005811; CAA06710.1; -.
DR   EMBL; AF121877; AAD30445.1; -.
DR   EMBL; AL079344; CAB45322.1; -.
DR   EMBL; AL161575; CAB79720.1; -.
DR   HSSP; P13652; 1CTT.
DR   InterPro; IPR002125; dCMP_cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
DR   PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW   Hydrolase.
SQ   SEQUENCE   337 AA;  37106 MW;  74E8CFF113EB46FB CRC64;

  Query Match             57.8%;  Score 37;  DB 10;  Length 337;
  Best Local Similarity   63.6%;  Pred. No. 61;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||||| ::
Db     186 CLTLSGSAGEI 196

RESULT   5
Q9SWZ4
ID   Q9SWZ4      PRELIMINARY;      PRT;   337 AA.
AC   Q9SWZ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CYTIDINE DEAMINASE 2 (EC 3.5.4.5).
GN   CDA2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Faivre-Nitschke S.E., Grienenberger J.M., Gualberto J.M.;
RT   "Cloning and characterisation of a cytidine deaminase gene family from
RT   Arabidopsis thaliana.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF080676; AAC69567.2; -.
DR   HSSP; P13652; 1CTT.
DR   InterPro; IPR002125; dCMP_cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
DR   PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW   Hydrolase.
SQ   SEQUENCE   337 AA;  37110 MW;  64ECCFF113EF46F8 CRC64;

  Query Match             57.8%;  Score 37;  DB 10;  Length 337;
  Best Local Similarity   63.6%;  Pred. No. 61;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||||| ::
Db     186 CLTLSGSAGEI 196

RESULT   6
Q91SH2
ID   Q91SH2      PRELIMINARY;      PRT;   552 AA.
AC   Q91SH2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GP35.
OS   Guinea pig cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Muromegalovirus.
OX   NCBI_TaxID=33706;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Biegalke B.J.;
RT   "Identification of a cluster of late genes in guinea pig
RT   cytomegalovirus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF355272; AAK43593.1; -.
SQ   SEQUENCE   552 AA;  63280 MW;  268404E1641D7C60 CRC64;

  Query Match             57.8%;  Score 37;  DB 12;  Length 552;
  Best Local Similarity   54.5%;  Pred. No. 1e+02;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           ||:: |:|:||
Db     442 CLMMEGAAQQL 452

RESULT   7
Q9IES5
ID   Q9IES5      PRELIMINARY;      PRT;   173 AA.
AC   Q9IES5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   19 KDA PROTEIN (FRAGMENT).
GN   P19.
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LETTUCE;
RA   Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,
RA   Duffus J.E., Koike S.T., Wisler G.C.;
RT   "Characterization of distinct tombusviruses that cause diseases of
RT   lettuce and tomato in the Western United States.";
RL   Phytopathology 91:797-806(2001).
DR   EMBL; AJ288927; CAC01093.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   173 AA;  19474 MW;  69244B482E45AFA9 CRC64;

  Query Match             56.2%;  Score 36;  DB 12;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130

RESULT   8
Q9IES9
ID   Q9IES9      PRELIMINARY;      PRT;   173 AA.
```

```
AC   Q9IES9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   19 KDA PROTEIN (FRAGMENT).
GN   P19.
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LETTUCE;
RA   Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,
RA   Duffus J.E., Koike S.T., Wisler G.C.;
RT   "Characterization of distinct tombusviruses that cause diseases of
RT   lettuce and tomato in the Western United States.";
RL   Phytopathology 91:797-806(2001).
DR   EMBL; AJ288919; CAC01089.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   173 AA;  19484 MW;  69244B482D8AAFA9 CRC64;

  Query Match             56.2%;  Score 36;  DB 12;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130


RESULT   9
Q9IES8
ID   Q9IES8      PRELIMINARY;      PRT;   173 AA.
AC   Q9IES8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   19 KDA PROTEIN (FRAGMENT).
GN   P19.
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LETTUCE;
RA   Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,
RA   Duffus J.E., Koike S.T., Wisler G.C.;
RT   "Characterization of distinct tombusviruses that cause diseases of
RT   lettuce and tomato in the Western United States.";
RL   Phytopathology 91:797-806(2001).
DR   EMBL; AJ288920; CAC01090.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   173 AA;  19484 MW;  69244B482D8AAFA9 CRC64;

  Query Match             56.2%;  Score 36;  DB 12;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130


RESULT  10
Q9IES7
ID   Q9IES7      PRELIMINARY;      PRT;   173 AA.
AC   Q9IES7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   19 KDA PROTEIN (FRAGMENT).
GN   P19.
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LETTUCE;
RA   Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,
RA   Duffus J.E., Koike S.T., Wisler G.C.;
RT   "Characterization of distinct tombusviruses that cause diseases of
RT   lettuce and tomato in the Western United States.";
RL   Phytopathology 91:797-806(2001).
DR   EMBL; AJ288921; CAC01091.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   173 AA;  19474 MW;  69244B482E45AFA9 CRC64;

  Query Match             56.2%;  Score 36;  DB 12;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130


RESULT  11
Q9IES6
ID   Q9IES6      PRELIMINARY;      PRT;   173 AA.
AC   Q9IES6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   19 KDA PROTEIN (FRAGMENT).
GN   P19.
OS   Cucumber necrosis virus (CNV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC   Tombusvirus.
OX   NCBI_TaxID=12143;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LETTUCE;
RA   Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,
RA   Duffus J.E., Koike S.T., Wisler G.C.;
RT   "Characterization of distinct tombusviruses that cause diseases of
RT   lettuce and tomato in the Western United States.";
RL   Phytopathology 91:797-806(2001).
DR   EMBL; AJ288922; CAC01092.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   173 AA;  19484 MW;  69244B482D8AAFA9 CRC64;

  Query Match             56.2%;  Score 36;  DB 12;  Length 173;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           || ||| :| |
Db     120 CLTLSGGSRTL 130


RESULT  12
Q9ATM2
ID   Q9ATM2      PRELIMINARY;      PRT;   243 AA.
AC   Q9ATM2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   SMALL BASIC MEMBRANE INTEGRAL PROTEIN ZMSIP1-2.
OS   Zea mays (Maize).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21140306; PubMed=11244102;
RA   Chaumont F., Barrieu F., Wojcik E., Chrispeels M.J., Jung R.;
RT   "Aquaporins constitute a large and highly divergent protein family in
RT   maize.";
RL   Plant Physiol. 125:1206-1215(2001).
DR   EMBL; AF326498; AAK26765.1; -.
DR   InterPro; IPR000425; MIP.
DR   Pfam; PF00230; MIP; 1.
DR   PRINTS; PR00783; MINTRINSICP.
SQ   SEQUENCE   243 AA;  25703 MW;  5B37A379977C8C9E CRC64;


  Query Match             56.2%;  Score 36;  DB 10;  Length 243;
  Best Local Similarity   87.5%;  Pred. No. 67;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSA 8
            ||||||:|
Db     175 CLVLSGAA 182


RESULT  13
Q9ZVD2
ID   Q9ZVD2      PRELIMINARY;      PRT;   260 AA.
AC   Q9ZVD2;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G27080 PROTEIN (AT2G27080/T20P8.13) (UNKNOWN PROTEIN).
GN   AT2G27080.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA   Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Chen H., Cheuk R.,
RA   Chung M.K., Kim C., Lin J., Liu S.X., Pham P.K., Sakano H., Shinn P.,
RA   Yamada K., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005623; AAC77866.1; -.
DR   EMBL; AY052291; AAK96484.1; -.
DR   EMBL; AF370591; AAK43910.1; -.
SQ   SEQUENCE   260 AA;  28975 MW;  36872A3E08F24371 CRC64;


  Query Match             56.2%;  Score 36;  DB 10;  Length 260;
  Best Local Similarity   81.8%;  Pred. No. 72;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLT 12
            ||||||  |||
Db     181 LVLSGSKIQLT 191


RESULT  14
Q9AC55
ID   Q9AC55      PRELIMINARY;      PRT;   262 AA.
AC   Q9AC55;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENOYL-COA HYDRATASE/ISOMERASE FAMILY PROTEIN.
GN   CC0006.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005675; AAK21994.1; -.
DR   HSSP; P14604; 2DUB.
DR   TIGR; CC0006; -.
DR   InterPro; IPR001753; Enoyl_CoA_hydrtse.
DR   Pfam; PF00378; ECH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Isomerase; Complete proteome.
SQ   SEQUENCE   262 AA;  27631 MW;  9CB076DD3F6B9D71 CRC64;


  Query Match             56.2%;  Score 36;  DB 16;  Length 262;
  Best Local Similarity   66.7%;  Pred. No. 73;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSAR 9
            |:||:|||:
Db      55 CIVLTGSAK 63


RESULT  15
P72593
ID   P72593      PRELIMINARY;      PRT;   315 AA.
AC   P72593;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
```

DE IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECB.
GN FECB OR SLR1319.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90899; BAA16593.1; -.
DR InterPro; IPR002491; Peripla_BP.
DR Pfam; PF01497; Peripla_BP_2; 1.
KW Complete proteome.
SQ SEQUENCE 315 AA; 34868 MW; 3F984C0C26352A41 CRC64;

Query Match 56.2%; Score 36; DB 16; Length 315;
Best Local Similarity 63.6%; Pred. No. 88;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 CLVLSGSARQL 11
           |||| | |:|:
Db      11 CLVLFGCAKQV 21
```

Search completed: June 6, 2002, 06:22:08
Job time: 377 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 52.04 Seconds
(without alignments)
27.747 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	64	100.0	13	19	AAW47590	T-cell receptor CD
2	45.5	71.1	267	19	AAW47588	T-cell receptor al
3	42.5	66.4	12	19	AAW47591	T-cell receptor CD
4	38	59.4	205	21	AAB56211	Human secreted pro
5	38	59.4	222	21	AAB56352	Human secreted pro
6	38	59.4	537	22	ABG26118	Novel human diagno
7	38	59.4	663	22	ABG08826	Novel human diagno
8	38	59.4	663	22	ABG29486	Novel human diagno
9	36	56.2	34	22	ABB32436	Peptide #5087 enco
10	36	56.2	34	22	AAM58353	Human brain expres
11	36	56.2	34	22	AAM18673	Peptide #5107 enco
12	35	54.7	147	22	AAU39475	Propionibacterium
13	35	54.7	284	22	AAG84910	Shrimp white spot
14	35	54.7	334	22	AAG91501	C glutamicum prote
15	35	54.7	334	22	AAB79276	Corynebacterium gl
16	35	54.7	386	22	ABB63279	Drosophila melanog
17	35	54.7	436	22	AAB82203	Megalomicin biosyn
18	35	54.7	2703	22	ABB63299	Drosophila melanog
19	34	53.1	28	16	AAR80130	TCR CDR3 V-alpha r
20	34	53.1	44	22	AAM69627	Human bone marrow
21	34	53.1	44	22	AAM17440	Peptide #3874 enco
22	34	53.1	55	22	ABB38722	Peptide #6228 enco
23	34	53.1	55	22	AAM59355	Human brain expres
24	34	53.1	55	22	AAM71910	Human bone marrow
25	34	53.1	55	22	AAM32191	Peptide #6228 enco
26	34	53.1	59	20	AAY13021	Human secreted pro
27	34	53.1	61	22	AAU53825	Propionibacterium
28	34	53.1	78	20	AAY36016	Extended human sec
29	34	53.1	99	22	AAU14591	Human novel protei
30	34	53.1	106	11	AAR06641	Histidinol dehydro
31	34	53.1	112	22	AAG74229	Human colon cancer
32	34	53.1	118	22	AAU14355	Human novel protei
33	34	53.1	119	21	AAG18823	Zea mays protein f
34	34	53.1	157	21	AAG18821	Zea mays protein f
35	34	53.1	174	22	AAU59698	Propionibacterium
36	34	53.1	182	21	AAG24465	Arabidopsis thalia
37	34	53.1	183	21	AAG24464	Arabidopsis thalia
38	34	53.1	183	21	AAG43167	Arabidopsis thalia
39	34	53.1	185	21	AAG24463	Arabidopsis thalia
40	34	53.1	185	21	AAG43166	Arabidopsis thalia
41	34	53.1	189	21	AAG43165	Arabidopsis thalia
42	34	53.1	269	19	AAW53952	Bacillus subtilis
43	34	53.1	271	21	AAB40344	Human ORFX ORF108
44	34	53.1	297	21	AAY76748	Human protein kina
45	34	53.1	297	22	AAE06206	Human protein kina

ALIGNMENTS

RESULT 1
AAW47590
ID AAW47590 standard; peptide; 13 AA.
XX
AC AAW47590;
XX
DT 26-JUN-1998 (first entry)
XX
DE T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3.
XX
OS Homo sapiens.
XX
PN DE19625191-A1.
XX
PD 02-JAN-1998.
XX
PF 24-JUN-1996; 96DE-1025191.
XX
PR 24-JUN-1996; 96DE-1025191.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Schendel D;
XX
DR WPI; 1998-053442/06.
DR N-PSDB; AAV18707.
XX
PT Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX

```
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   13 AA;


  Query Match           100.0%;  Score 64;  DB 19;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 6.2e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          |||||||||||||
Db      1 clvlsgsarqltf 13


RESULT   2
AAW47588
ID   AAW47588 standard; Protein; 267 AA.
XX
AC   AAW47588;
XX
DT   26-JUN-1998  (first entry)
XX
DE   T-cell receptor alpha-chain.
XX
KW   Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma.
XX
OS   Homo sapiens.
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   N-PSDB; AAV18705.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Pages 11-13; 30pp; German.
XX
CC   The present sequence is the alpha-chain of a human T-cell
CC   receptor (TCR), which can be used in the diagnosis, monitoring,
CC   prevention and therapy of a tumour disease, specifically renal
CC   cell carcinoma.
XX
SQ   Sequence   267 AA;


  Query Match           71.1%;  Score 45.5;  DB 19;  Length 267;
  Best Local Similarity  84.6%;  Pred. No. 0.93;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 CLVLSGSARQLTF 13
          |||  ||||||||
Db    105 clv-ggsarqltf 116


RESULT   3
AAW47591
ID   AAW47591 standard; peptide; 12 AA.
XX
AC   AAW47591;
XX
DT   26-JUN-1998  (first entry)
XX
DE   T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3.
XX
OS   Homo sapiens.
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   N-PSDB; AAV18708.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   12 AA;


  Query Match           66.4%;  Score 42.5;  DB 19;  Length 12;
  Best Local Similarity  76.9%;  Pred. No. 0.1;
  Matches   10;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 CLVLSGSARQLTF 13
          ||  :||||||||
Db      1 cla-tgsarqltf 12


RESULT   4
AAB56211
ID   AAB56211 standard; Protein; 205 AA.
XX
AC   AAB56211;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human secreted protein sequence encoded by gene 135 SEQ ID NO:305.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; pathological condition;
KW   autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW   neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW   cerebrovascular disorder; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; skin aging; food additive; preservative.
XX
OS   Homo sapiens.
XX
```

PN WO200070042-A1.
XX
PD 23-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-US12788.
XX
PR 13-MAY-1999; 99US-0134068.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX
DR WPI; 2000-679828/66.
DR N-PSDB; AAC99952.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 11; Page 995-996; 1065pp; English.
XX
CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and opthalmalogical. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention.
XX
SQ Sequence 205 AA;

Query Match 59.4%; Score 38; DB 21; Length 205;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQLTF 13
||:| ||:| :|
Db 88 clllcgssrttsf 100

RESULT 5
AAB56352
ID AAB56352 standard; Protein; 222 AA.
XX
AC AAB56352;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 135 SEQ ID NO:446.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative.
XX
OS Homo sapiens.
XX
PN WO200070042-A1.
XX
PD 23-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-US12788.
XX
PR 13-MAY-1999; 99US-0134068.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX
DR WPI; 2000-679828/66.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; Page 1057; 1065pp; English.
XX
CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and opthalmalogical. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention.
XX
SQ Sequence 222 AA;

Query Match 59.4%; Score 38; DB 21; Length 222;
Best Local Similarity 53.8%; Pred. No. 23;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQLTF 13
||:| ||:| :|
Db 105 clllcgssrttsf 117

RESULT 6
ABG26118
ID ABG26118 standard; Protein; 537 AA.
XX
AC ABG26118;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26109.

```
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS90305.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 56477; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   537 AA;


  Query Match             59.4%;  Score 38;  DB 22;  Length 537;
  Best Local Similarity   72.7%;  Pred. No. 63;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           ||| | ||||:
Db      55 clvystsarqi 65


RESULT   7
ABG08826
ID   ABG08826 standard; Protein; 663 AA.
XX
AC   ABG08826;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #8817.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS73013.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 39185; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   663 AA;


  Query Match             59.4%;  Score 38;  DB 22;  Length 663;
  Best Local Similarity   72.7%;  Pred. No. 80;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQL 11
           ||| | ||||:
Db      55 clvystsarqi 65


RESULT   8
ABG29486
ID   ABG29486 standard; Protein; 663 AA.
XX
AC   ABG29486;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29477.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS93673.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 59845; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 663 AA;

Query Match 59.4%; Score 38; DB 22; Length 663;
Best Local Similarity 72.7%; Pred. No. 80;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQL 11
|||| | ||||:
Db 55 clvystsarqi 65

RESULT 9
ABB32436
ID ABB32436 standard; Peptide; 34 AA.
XX
AC ABB32436;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #5087 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27; SEQ ID NO 15404; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 34 AA;

Query Match 56.2%; Score 36; DB 22; Length 34;
Best Local Similarity 58.3%; Pred. No. 6.4;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CLVLSGSARQLT 12
|||| | | |:
Db 2 clvleggkrdls 13

RESULT 10
AAM58353
ID AAM58353 standard; Protein; 34 AA.
XX
AC AAM58353;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30458.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

```
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains
XX
PS   Example 4; SEQ ID NO: 30458; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
SQ   Sequence   34 AA;


  Query Match             56.2%;  Score 36;  DB 22;  Length 34;
  Best Local Similarity   58.3%;  Pred. No. 6.4;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           |||| |  | |:
Db       2 clvleggkrdls 13


RESULT  11
AAM18673
ID   AAM18673 standard; Protein; 34 AA.
XX
AC   AAM18673;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #5107 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 23499; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   34 AA;


  Query Match             56.2%;  Score 36;  DB 22;  Length 34;
  Best Local Similarity   58.3%;  Pred. No. 6.4;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           |||| |  | |:
Db       2 clvleggkrdls 13


RESULT  12
AAU39475
ID   AAU39475 standard; Protein; 147 AA.
XX
AC   AAU39475;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #371.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
```

```
DR   N-PSDB; AAS59507.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 670; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   147 AA;


  Query Match             54.7%;  Score 35;  DB 22;  Length 147;
  Best Local Similarity   63.6%;  Pred. No. 55;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLT 12
           ||::|:| |||
Db     107 lvvtgtahqlt 117


RESULT  13
AAG84910
ID   AAG84910 standard; Protein; 284 AA.
XX
AC   AAG84910;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) protein 1.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp.
XX
OS   White spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US28888.
XX
PR   24-NOV-1999;   99CN-0124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   N-PSDB; AAH62690.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection -
XX
PS   Claim 1; Figure 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp.
XX
SQ   Sequence   284 AA;


  Query Match             54.7%;  Score 35;  DB 22;  Length 284;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 LSGSARQLTF 13
           :|| :|||| :
Db     199 isgksrqlty 208


RESULT  14
AAG91501
ID   AAG91501 standard; Protein; 334 AA.
XX
AC   AAG91501;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum protein fragment SEQ ID NO: 5255.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   N-PSDB; AAH66720.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 17; SEQ ID NO: 5255; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
```

CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 334 AA;

Query Match 54.7%; Score 35; DB 22; Length 334;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLSGSARQLTF 13
|::||| |||
Db 237 vitgsatdltf 247

RESULT 15
AAB79276
ID AAB79276 standard; Protein; 334 AA.
XX
AC AAB79276;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:68.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
DR N-PSDB; AAF71393.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 20; Page 238; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 334 AA;

Query Match 54.7%; Score 35; DB 22; Length 334;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLSGSARQLTF 13
|::||| |||
Db 237 vitgsatdltf 247

Search completed: June 6, 2002, 06:17:20
Job time: 124 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 21.84 Seconds
(without alignments)
14.539 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	53.1	28	3	US-08-963-121C-11	Sequence 11, Appl
2	34	53.1	28	4	US-09-543-513-11	Sequence 11, Appl
3	34	53.1	28	5	PCT-US95-04803-12	Sequence 12, Appl
4	34	53.1	59	4	US-08-905-223-305	Sequence 305, App
5	34	53.1	106	6	5204252-6	Patent No. 5204252
6	34	53.1	269	3	US-08-926-842B-59	Sequence 59, Appl
7	34	53.1	297	3	US-09-173-581-1	Sequence 1, Appli
8	34	53.1	297	4	US-09-420-915-1	Sequence 1, Appli
9	33.5	52.3	115	2	US-08-485-449-4	Sequence 4, Appli
10	33.5	52.3	376	2	US-08-485-449-5	Sequence 5, Appli
11	33.5	52.3	389	2	US-08-485-449-2	Sequence 2, Appli
12	33.5	52.3	389	2	US-08-485-449-6	Sequence 6, Appli
13	33.5	52.3	389	2	US-08-485-449-7	Sequence 7, Appli
14	33	51.6	376	3	US-09-025-691-3	Sequence 3, Appli
15	33	51.6	762	4	US-09-228-986-114	Sequence 114, App
16	32.5	50.8	1500	4	US-09-323-472A-2	Sequence 2, Appli
17	32.5	50.8	1500	4	US-09-323-472A-4	Sequence 4, Appli
18	32.5	50.8	1500	4	US-09-323-472A-12	Sequence 12, Appl
19	32.5	50.8	1500	4	US-09-323-472A-14	Sequence 14, Appl
20	32	50.0	252	4	US-09-199-637A-176	Sequence 176, App
21	31	48.4	140	1	US-08-664-596B-2	Sequence 2, Appli
22	31	48.4	140	2	US-08-738-367-2	Sequence 2, Appli
23	31	48.4	165	3	US-08-685-808-4	Sequence 4, Appli
24	31	48.4	165	4	US-08-505-860C-4	Sequence 4, Appli
25	31	48.4	202	1	US-08-469-486-56	Sequence 56, Appl
26	31	48.4	202	2	US-08-469-658-56	Sequence 56, Appl
27	31	48.4	240	2	US-08-738-367-7	Sequence 7, Appli
28	31	48.4	255	1	US-08-459-264-4	Sequence 4, Appli
29	31	48.4	255	1	US-08-459-263-4	Sequence 4, Appli
30	31	48.4	807	4	US-08-855-910-9	Sequence 9, Appli
31	31	48.4	824	1	US-08-221-750A-3	Sequence 3, Appli
32	31	48.4	844	3	US-09-029-267-20	Sequence 20, Appl
33	31	48.4	890	1	US-08-472-934-8	Sequence 8, Appli
34	31	48.4	890	2	US-08-323-460A-8	Sequence 8, Appli
35	31	48.4	890	2	US-08-461-146C-8	Sequence 8, Appli
36	31	48.4	890	3	US-08-461-145C-8	Sequence 8, Appli
37	31	48.4	890	4	US-08-628-829-12	Sequence 12, Appl
38	31	48.4	1247	1	US-08-472-934-10	Sequence 10, Appl
39	31	48.4	1247	2	US-08-323-460A-10	Sequence 10, Appl
40	31	48.4	1247	2	US-08-461-146C-10	Sequence 10, Appl
41	31	48.4	1247	3	US-08-461-145C-10	Sequence 10, Appl
42	31	48.4	1597	4	US-09-423-890-13	Sequence 13, Appl
43	31	48.4	1597	4	US-08-628-829-14	Sequence 14, Appl
44	30	46.9	107	1	US-08-276-852-107	Sequence 107, App
45	30	46.9	107	1	US-08-899-575-107	Sequence 107, App

ALIGNMENTS

```
RESULT 1
US-08-963-121C-11
; Sequence 11, Application US/08963121C
; Patent No. 6084087
; GENERAL INFORMATION:
;   APPLICANT: Friedman, Steven M
;   APPLICANT: Crow, Mary K
;   APPLICANT: Yi, Y.
;   APPLICANT: Tumang, Joseph
;   APPLICANT: Sun, Guang-Rong
;   TITLE OF INVENTION: Conserved T-Cell Receptor Sequences
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Darby & Darby PC
;     STREET: 805 Third Avenue
;     CITY: New York
;     STATE: New York
;     COUNTRY: US
;     ZIP: 10022
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/963,121C
;     FILING DATE: October 28, 1997
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/427,009
;     FILING DATE: April, 24, 1995
;     APPLICATION NUMBER: 08/229,285
;     FILING DATE: April 18, 1994
;     APPLICATION NUMBER: 07/766,751, Patent No. 6084087 5,480,895
;     FILING DATE: September 27, 1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Ludwig, S. Peter
;     REGISTRATION NUMBER: 25,351
;     REFERENCE/DOCKET NUMBER: 5983/17499-US1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-527-7700
;     TELEFAX: 212-753-6237
;     TELEX: 236687
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 28 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
```

```
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Va3.1-aa
US-08-963-121C-11


  Query Match             53.1%;  Score 34;  DB 3;  Length 28;
  Best Local Similarity   53.8%;  Pred. No. 3.1;
  Matches    7;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          |  | ||  :|||
Db      1 CATLGGSNYKLTF 13


RESULT   2
US-09-543-513-11
; Sequence 11, Application US/09543513
; Patent No. 6303750
;  GENERAL INFORMATION:
;    APPLICANT:  Friedman, Steven M
;    APPLICANT:  Crow, Mary K
;    APPLICANT:  Yi, Y.
;    APPLICANT:  Tumang, Joseph
;    APPLICANT:  Sun, Guang-Rong
;    TITLE OF INVENTION:  Conserved T-Cell Receptor Sequences
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/543,513
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/963,121
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  5983/17499-US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Va3.1-aa
US-09-543-513-11


  Query Match             53.1%;  Score 34;  DB 4;  Length 28;
  Best Local Similarity   53.8%;  Pred. No. 3.1;
  Matches    7;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          |  | ||  :|||
Db      1 CATLGGSNYKLTF 13


RESULT   3
PCT-US95-04803-12
; Sequence 12, Application PC/TUS9504803
;  GENERAL INFORMATION:
;    APPLICANT:  New York Society For the ruptured and
;    APPLICANT:  Crippled Maintaining The Hospital for
;    APPLICANT:  Special Surgery
;    APPLICANT:  INVENTORS: Friedman, Steven M
;    APPLICANT:  Crow, Mary K
;    APPLICANT:  Yi, Y.
;    APPLICANT:  Tumang, Joseph
;    APPLICANT:  Sun, Guang-Rong
;    TITLE OF INVENTION:  Conserved T-Cell Receptor Sequences
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04803
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  5983/09449
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Va3.1-aa
PCT-US95-04803-12


  Query Match             53.1%;  Score 34;  DB 5;  Length 28;
  Best Local Similarity   53.8%;  Pred. No. 3.1;
  Matches    7;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          |  | ||  :|||
Db      1 CATLGGSNYKLTF 13


RESULT   4
US-08-905-223-305
; Sequence 305, Application US/08905223
; Patent No. 6222029
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  305:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  59 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  -21..-1
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 6.9
;      OTHER INFORMATION:  seq LLACGSLLPGLWQ/HL
US-08-905-223-305


  Query Match             53.1%;  Score 34;  DB 4;  Length 59;
  Best Local Similarity   50.0%;  Pred. No. 7.3;
  Matches    6;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           | | |||: |::
Db      36 CTVCSGSSEQIS 47


RESULT   5
5204252-6
;Patent No. 5204252
;    APPLICANT: CREGG,JAMES M.;GLEESON,MARTIN A.;HAAS,LISA
;PICATAGGIO,STEPHEN
;    TITLE OF INVENTION: CANDIDA TROPICALIS TRANSFORMATION SYSTEM
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/386,837
;      FILING DATE: 27-JUL-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 308,481
;      FILING DATE: 08-FEB-1989
;SEQ ID NO:6:
;      LENGTH: 106
5204252-6


  Query Match             53.1%;  Score 34;  DB 6;  Length 106;
  Best Local Similarity   41.7%;  Pred. No. 14;
  Matches    5;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLT 12
           |:|:::| |: :|
Db      20 CIVMAGGAQAVT 31


RESULT   6
US-08-926-842B-59
; Sequence 59, Application US/08926842B
; Patent No. 6030807
;  GENERAL INFORMATION:
;    APPLICANT:  Sa-No. 6030807ueira, Isabel
;    APPLICANT:  de Lencastre, Herminia
;    TITLE OF INVENTION:  HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/926,842B
;      FILING DATE:  10-SEP-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-089 N
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus subtilis
;    FEATURE:
;      OTHER INFORMATION:  /product= "araL"
US-08-926-842B-59


  Query Match             53.1%;  Score 34;  DB 3;  Length 269;
  Best Local Similarity   77.8%;  Pred. No. 41;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQ 10
           |||:|||:|
Db     233 LVLTGSAKQ 241


RESULT   7
US-09-173-581-1
; Sequence 1, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
```

```
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/173,581A
;  CURRENT FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 297
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 119819
US-09-173-581-1


  Query Match             53.1%;  Score 34;  DB 3;  Length 297;
  Best Local Similarity   60.0%;  Pred. No. 46;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQ 10
           | :||||: |
Db      93 CSILSGSSNQ 102


RESULT   8
US-09-420-915-1
; Sequence 1, Application US/09420915
; Patent No. 6264947
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/420,915
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/173,581
;  EARLIER FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 297
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 119819
US-09-420-915-1


  Query Match             53.1%;  Score 34;  DB 4;  Length 297;
  Best Local Similarity   60.0%;  Pred. No. 46;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQ 10
           | :||||: |
Db      93 CSILSGSSNQ 102


RESULT   9
US-08-485-449-4
; Sequence 4, Application US/08485449
; Patent No. 5824789
;  GENERAL INFORMATION:
;    APPLICANT:  VANDENBERG, DAVID
;    TITLE OF INVENTION:  HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,449
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20296-20035.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  115 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-485-449-4


  Query Match             52.3%;  Score 33.5;  DB 2;  Length 115;
  Best Local Similarity   75.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSG-SARQL 11
           || ||| | |||
Db      48 CLTLSGLSKRQL 59


RESULT  10
US-08-485-449-5
; Sequence 5, Application US/08485449
; Patent No. 5824789
;  GENERAL INFORMATION:
;    APPLICANT:  VANDENBERG, DAVID
;    TITLE OF INVENTION:  HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/485,449
;       FILING DATE:
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KONSKI, ANTOINETTE F.
;       REGISTRATION NUMBER:  34,202
;       REFERENCE/DOCKET NUMBER:  20296-20035.00
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 813-5600
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  376 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-485-449-5


  Query Match           52.3%;  Score 33.5;  DB 2;  Length 376;
  Best Local Similarity  75.0%;  Pred. No. 76;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSG-SARQL 11
           || ||| | |||
Db      44 CLTLSGLSKRQL 55


RESULT  11
US-08-485-449-2
; Sequence 2, Application US/08485449
; Patent No. 5824789
;  GENERAL INFORMATION:
;    APPLICANT:  VANDENBERG, DAVID
;    TITLE OF INVENTION:  HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,449
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20296-20035.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  389 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-485-449-2


  Query Match           52.3%;  Score 33.5;  DB 2;  Length 389;
  Best Local Similarity  75.0%;  Pred. No. 79;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSG-SARQL 11
           || ||| | |||
Db      48 CLTLSGLSKRQL 59


RESULT  12
US-08-485-449-6
; Sequence 6, Application US/08485449
; Patent No. 5824789
;  GENERAL INFORMATION:
;    APPLICANT:  VANDENBERG, DAVID
;    TITLE OF INVENTION:  HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,449
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20296-20035.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  389 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-485-449-6


  Query Match           52.3%;  Score 33.5;  DB 2;  Length 389;
  Best Local Similarity  75.0%;  Pred. No. 79;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSG-SARQL 11
           || ||| | |||
Db      48 CLTLSGLSKRQL 59


RESULT  13
US-08-485-449-7
; Sequence 7, Application US/08485449
; Patent No. 5824789
;  GENERAL INFORMATION:
;    APPLICANT:  VANDENBERG, DAVID
;    TITLE OF INVENTION:  HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
```

```
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,449
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20296-20035.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  389 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-485-449-7


  Query Match             52.3%;  Score 33.5;  DB 2;  Length 389;
  Best Local Similarity   75.0%;  Pred. No. 79;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLVLSG-SARQL 11
           || ||| | |||
Db      48 CLTLSGLSKRQL 59


RESULT  14
US-09-025-691-3
; Sequence 3, Application US/09025691
; Patent No. 6069299
;  GENERAL INFORMATION:
;    APPLICANT:  Broadway, Roxanne M.
;    APPLICANT:  Harman, Gary E.
;    TITLE OF INVENTION:  FUNGUS AND INSECT CONTROL WITH
;    TITLE OF INVENTION:  CHITINOLYTIC ENZYMES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,691
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/20120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  376 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-025-691-3


  Query Match             51.6%;  Score 33;  DB 3;  Length 376;
  Best Local Similarity   63.6%;  Pred. No. 95;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLT 12
           || |||| ::|
Db      30 LVTSGSAEKIT 40


RESULT  15
US-09-228-986-114
; Sequence 114, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
;  APPLICANT: Strabala, Timothy
;  APPLICANT: Nieuwenhuizen, Niels
;  TITLE OF INVENTION: Compositions Isolated from Plant Cells
;  TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
;  FILE REFERENCE: 11000/1020
;  CURRENT APPLICATION NUMBER: US/09/228,986
;  CURRENT FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 130
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 114
;   LENGTH: 762
;   TYPE: PRT
;   ORGANISM: Pinus radiata
US-09-228-986-114


  Query Match             51.6%;  Score 33;  DB 4;  Length 762;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQLTF 13
           |: |||| || |
Db     673 LLQSGSAFQLVF 684


Search completed: June  6, 2002, 06:15:47
Job time: 31 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:21:22 ; Search time 14.13 Seconds
(without alignments)
88.405 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1824

Minimum DB seq length: 0
Maximum DB seq length: 13

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26.5	41.4	13	2	S57571	T cell receptor al
2	26	40.6	13	2	S47361	T-cell antigen rec
3	22	34.4	10	2	S23371	T-cell receptor al
4	22	34.4	13	2	PH1772	T cell receptor al
5	21	32.8	10	2	PT0212	T-cell receptor al
6	20	31.2	8	2	S08996	hypertrehalosemic
7	20	31.2	8	2	B49823	adipokinetic hormo
8	20	31.2	8	2	B44960	neuropeptide Led-C
9	20	31.2	8	2	A33995	adipokinetic hormo
10	20	31.2	8	2	S11545	adipokinetic hormo
11	20	31.2	9	2	A24244	adipokinetic hormo
12	20	31.2	10	2	JC1416	hypertrehalosemic
13	20	31.2	10	2	S09138	hypertrehalosemic
14	20	31.2	10	2	A31571	hypertrehalosemic/
15	20	31.2	10	2	B33995	hypotrehalosemic h
16	20	31.2	13	2	S32473	lymnaDFamide 3 - g
17	20	31.2	13	2	B56864	dipeptidyl-peptida
18	19	29.7	6	2	PT0568	T-cell receptor be
19	19	29.7	10	2	S48182	bacterioferritin -
20	19	29.7	10	2	A47364	placental lactogen
21	19	29.7	11	2	S23364	T-cell receptor al
22	19	29.7	11	2	S61797	T-cell-specific tr
23	19	29.7	12	2	A37933	Ig lambda chain J
24	18	28.1	6	2	PT0280	Ig heavy chain CRD
25	18	28.1	10	2	G60787	sperm-activating p
26	18	28.1	10	2	E60787	sperm-activating p
27	18	28.1	10	2	E60589	sperm-activating p
28	18	28.1	10	2	S70251	nitrogenase (EC 1.
29	18	28.1	10	2	C41946	T-cell receptor ga
30	18	28.1	11	2	PH1583	Ig H chain V-D-J r
31	18	28.1	11	2	PH0903	T-cell receptor be
32	17	26.6	9	2	S59902	glutathione transf
33	17	26.6	9	2	I50633	c-rel protein - ch
34	17	26.6	10	2	S65728	hemoglobin, extrac
35	17	26.6	12	2	PH1188	T-cell receptor al
36	17	26.6	13	2	B61233	conceptus protein
37	17	26.6	13	2	PH0787	T-cell receptor al
38	16	25.0	6	2	A60494	antineoplastic gly
39	16	25.0	7	2	PN0150	omega-gliadine 1'
40	16	25.0	8	2	PT0639	T-cell receptor be
41	16	25.0	9	2	PS0253	glycine cleavage s
42	16	25.0	9	2	A29477	diuretic neuropept
43	16	25.0	10	2	JN0025	mosact - sea urchi
44	16	25.0	11	2	F33098	214K exoantigen (v
45	16	25.0	12	2	C58502	58K bile and gallb

ALIGNMENTS

RESULT 1
S57571
T cell receptor alpha chain V-J region (clone PP7 and others) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 19-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 05-Nov-1999
C;Accession: S57571; S57573; S57576
R;Burrows, S.R.; Silins, S.L.; Moss, D.J.; Khanna, R.; Misko, I.S.; Argaet, V.P.
submitted to the EMBL Data Library, June 1995
A;Description: T cell receptor repertoire for a viral epitope in humans is diversifie
A;Reference number: S57494
A;Accession: S57571
A;Molecule type: mRNA
A;Residues: 1-13 <BUR>
A;Cross-references: EMBL:Z49948; NID:g887496; PIDN:CAA90219.1; PID:g887497
A;Experimental source: clone PP7
A;Accession: S57573
A;Molecule type: mRNA
A;Residues: 1-13 <BUL>
A;Cross-references: EMBL:Z49950; NID:g887500; PIDN:CAA90221.1; PID:g887501
A;Experimental source: clone TF1
A;Accession: S57576
A;Molecule type: mRNA
A;Residues: 1-13 <BUW>
A;Cross-references: EMBL:Z49952; NID:g887512; PIDN:CAA90223.1; PID:g887513
A;Experimental source: clone RL16
C;Keywords: T-cell receptor

Query Match 41.4%; Score 26.5; DB 2; Length 13;
Best Local Similarity 53.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

```
Qy      1 CLVLSGSARQLTF 13
          | || |:  :|||
Db      1 CAVLFGN-EKLTF 12
```

RESULT 2
S47361
T-cell antigen receptor VJ junction beta chain - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: S47361
R;Lehner, P.J.
submitted to the EMBL Data Library, August 1994
A;Description: Human HLA-A0201 restricted recognition of influenza A is dominated by
A;Reference number: S47355
A;Accession: S47361
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-13 <LEH>
A;Cross-references: EMBL:Z35685; NID:g527459; PIDN:CAA84754.1; PID:g527460

```
C;Keywords: T-cell receptor


  Query Match             40.6%;  Score 26;  DB 2;  Length 13;
  Best Local Similarity   46.2%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSARQLTF 13
           | || ||  :  |
Db       1 CSVLQGSPYEQYF 13


RESULT   3
S23371
T-cell receptor alpha chain J region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
C;Accession: S23371
R;Pluschke, G.; Ricken, G.; Taube, H.; Kroninger, S.; Melchers, I.; Peter, H.H.; Eichman
Eur. J. Immunol. 21, 2749-2754, 1991
A;Title: Biased T cell receptor V(alpha) region repertoire in the synovial fluid of rheu
A;Reference number: S23364; MUID:92037820
A;Accession: S23371
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-10 <PLU>
A;Cross-references: EMBL:X58166
C;Keywords: T-cell receptor


  Query Match             34.4%;  Score 22;  DB 2;  Length 10;
  Best Local Similarity   55.6%;  Pred. No. 6.3e+02;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 SGSARQLTF 13
           || | : ||
Db       2 SGEAGKSTF 10


RESULT   4
PH1772
T cell receptor alpha chain V region (clone 2V alpha 23-3) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: PH1772
R;Porcelli, S.; Yockey, C.E.; Brenner, M.B.; Balk, S.P.
J. Exp. Med. 178, 1-16, 1993
A;Title: Analysis of T cell antigen receptor (TCR) expression by human peripheral blood
.
A;Reference number: PH1754; MUID:93301585
A;Accession: PH1772
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-13 <POR>


  Query Match             34.4%;  Score 22;  DB 2;  Length 13;
  Best Local Similarity   83.3%;  Pred. No. 8.2e+02;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 ARQLTF 13
           | ||||
Db       7 AGQLTF 12


RESULT   5
PT0212
T-cell receptor alpha chain V-J region (4-1-E.2) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 30-May-1997
C;Accession: PT0212
R;Nakano, N.; Kikutani, H.; Nishimoto, H.; Kishimoto, T.
J. Exp. Med. 173, 1091-1097, 1991
A;Title: T cell receptor V gene usage of islet beta cell-reactive T cells is not rest
A;Reference number: PT0209; MUID:91217621
A;Accession: PT0212
A;Molecule type: mRNA
A;Residues: 1-10 <NAK>
C;Keywords: T-cell receptor


  Query Match             32.8%;  Score 21;  DB 2;  Length 10;
  Best Local Similarity   44.4%;  Pred. No. 9.9e+02;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CLVLSGSAR 9
           | |  |: |
Db       1 CAVAGGADR 9


RESULT   6
S08996
hypertrehalosemic hormone II - oriental cockroach
N;Alternate names: Pea-CAH-II
C;Species: Blatta orientalis (oriental cockroach)
C;Date: 30-Jun-1992 #sequence_revision 24-Oct-1997 #text_change 31-Oct-1997
C;Accession: S08996
R;Gaede, G.; Rinehart, K.L.
Biol. Chem. Hoppe-Seyler 371, 345-354, 1990
A;Title: Primary structures of hypertrehalosaemic neuropeptides isolated from the cor
entalis and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bomb
A;Reference number: S08995; MUID:90253659
A;Accession: S08996
A;Molecule type: protein
A;Residues: 1-8 <GAE>
A;Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we h
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental


  Query Match             31.2%;  Score 20;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   7
B49823
adipokinetic hormone II - American cockroach
N;Alternate names: neuropeptide M-II; periplanetin CC-1
C;Species: Periplaneta americana (American cockroach)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: B49823; A05170
R;Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller,
Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984
A;Title: Isolation and primary structure of two peptides with cardioacceleratory and
A;Reference number: A49823; MUID:84298179
A;Accession: B49823
A;Molecule type: protein
A;Residues: 1-8 <SCA>
R;Witten, J.L.; Schaffer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr.,
Biochem. Biophys. Res. Commun. 124, 350-358, 1984
A;Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment
A;Reference number: A90118; MUID:85046530
A;Accession: A05170
A;Molecule type: protein
A;Residues: 'E',2-8 <WIT>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental
```

Query Match 31.2%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 QLTF 13
           ||||
Db       1 QLTF 4
```

RESULT 8
B44960
neuropeptide Led-CC-II - Colorado potato beetle
C;Species: Leptinotarsa decemlineata (Colorado potato beetle)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: B44960
R;Gaede, G.; Kellner, R.
Peptides 10, 1287-1289, 1989
A;Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and
A;Reference number: A44960; MUID:90160053
A;Accession: B44960
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 QLTF 13
           ||||
Db       1 QLTF 4
```

RESULT 9
A33995
adipokinetic hormone - black horse fly
C;Species: Tabanus atratus (black horse fly)
C;Date: 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change 31-Oct-1997
C;Accession: A33995
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhang,
Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989
A;Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotrehal
A;Reference number: A33995; MUID:90046758
A;Accession: A33995
A;Molecule type: protein
A;Residues: 1-8 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted
F;8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 31.2%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 QLTF 13
           ||||
Db       1 QLTF 4
```

RESULT 10
S11545
adipokinetic hormone - nestling-sucking blowfly
C;Species: Protophormia terraenovae (nestling-sucking blowfly)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 31-Oct-1997
C;Accession: S11545
R;Gaede, G.; Wilps, H.; Kellner, R.
Biochem. J. 269, 309-313, 1990
A;Title: Isolation and structure of a novel charged member of the red-pigment-concent
erraenovae (Diptera).
A;Reference number: S11545; MUID:90351345
A;Accession: S11545
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 QLTF 13
           ||||
Db       1 QLTF 4
```

RESULT 11
A24244
adipokinetic hormone - bollworm
N;Alternate names: Hez-AKH
C;Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)
C;Date: 31-Mar-1988 #sequence_revision 23-Mar-1995 #text_change 31-Oct-1997
C;Accession: A24244
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridg
Biochem. Biophys. Res. Commun. 135, 622-628, 1986
A;Title: Isolation and primary structure of a peptide from the corpora cardiaca of He
A;Reference number: A24244; MUID:86186794
A;Accession: A24244
A;Molecule type: protein
A;Residues: 1-9 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 QLTF 13
           ||||
Db       1 QLTF 4
```

RESULT 12
JC1416
hypertrehalosemic hormone I - stick insect (Carausius morosus)
N;Alternate names: neuropeptide Cam-HrTH-I
N;Contains: hypertrehalosemic factor II
C;Species: Carausius morosus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: JC1416; S07157
R;Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.
Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992
A;Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick
A;Reference number: JC1416; MUID:93129188
A;Accession: JC1416
A;Molecule type: protein
A;Residues: 1-10 <GAE1>
R;Gaede, G.; Rinehart Jr., K.L.
Biol. Chem. Hoppe-Seyler 368, 67-75, 1987
A;Title: Primary structure of the hypertrehalosaemic factor II from the corpus cardia
A;Reference number: S07157; MUID:87157103
A;Accession: S07157
A;Molecule type: protein

```
A;Residues: 'Z',2-10 <GAE2>
C;Comment: Hypertrehalosemic factor II lacks the tryptophan modification.
C;Comment: This peptide raises hemolymph levels of trehalose in the cockroach Periplanet
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide; pyroglutamic aci
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Binding site: carbohydrate (Trp) (covalent) #status experimental
F;10/Modified site: amidated carboxyl end (Thr) #status experimental


  Query Match             31.2%;  Score 20;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT  13
S09138
hypertrehalosemic hormone II - stick insect (Extatosoma tiaratum)
N;Alternate names: Cam-HrTH-II
C;Species: Extatosoma tiaratum
C;Date: 30-Jun-1992 #sequence_revision 24-Oct-1997 #text_change 31-Oct-1997
C;Accession: S09138
R;Gaede, G.; Rinehart, K.L.
Biol. Chem. Hoppe-Seyler 371, 345-354, 1990
A;Title: Primary structures of hypertrehalosaemic neuropeptides isolated from the corpor
entalis and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombard
A;Reference number: S08995; MUID:90253659
A;Accession: S09138
A;Molecule type: protein
A;Residues: 1-10 <GAE>
A;Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;10/Modified site: amidated carboxyl end (Thr) #status experimental


  Query Match             31.2%;  Score 20;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT  14
A31571
hypertrehalosemic/adipokinetic hormone - bollworm
N;Alternate names: Hez-HrTH
C;Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)
C;Date: 30-Jun-1989 #sequence_revision 23-Mar-1995 #text_change 31-Oct-1997
C;Accession: A31571
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Bird, T.G.; Tseng, C.M.; Zhang, Y.S
Biochem. Biophys. Res. Commun. 155, 344-350, 1988
A;Title: Isolation and primary structure of a neuropeptide hormone from Heliothis zea wi
A;Reference number: A31571; MUID:88326324
A;Accession: A31571
A;Molecule type: protein
A;Residues: 1-10 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;10/Modified site: amidated carboxyl end (Asn) #status experimental


  Query Match             31.2%;  Score 20;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT  15
B33995
hypotrehalosemic hormone - black horse fly
C;Species: Tabanus atratus (black horse fly)
C;Date: 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change 31-Oct-1997
C;Accession: B33995
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhan
Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989
A;Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotre
A;Reference number: A33995; MUID:90046758
A;Accession: B33995
A;Molecule type: protein
A;Residues: 1-10 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted
F;10/Modified site: amidated carboxyl end (Tyr) #status predicted


  Query Match             31.2%;  Score 20;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


Search completed: June  6, 2002, 06:23:43
Job time: 141 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:27 ; Search time 10.15 Seconds
(without alignments)
49.592 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 556

Minimum DB seq length: 0
Maximum DB seq length: 13

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	35.9	11	1	ES1_RAT	P56571 rattus norv
2	20	31.2	8	1	AKH_TABAT	P14595 tabanus atr
3	20	31.2	8	1	HTF2_PERAM	P04549 periplaneta
4	20	31.2	10	1	HTF2_CARMO	P11385 carausius m
5	20	31.2	10	1	HTF_HELZE	P16353 heliothis z
6	20	31.2	10	1	HTF_TABAT	P14596 tabanus atr
7	20	31.2	13	1	NP3_LYMST	P80180 lymnaea sta
8	19	29.7	8	1	CAD1_ENTFA	P13268 enterococcu
9	18	28.1	13	1	ITB5_BOVIN	P80747 bos taurus
10	18	28.1	13	1	TEMA_RANTE	P56917 rana tempor
11	18	28.1	13	1	TEMF_RANTE	P56921 rana tempor
12	17	26.6	10	1	AKHX_LOCMI	P81626 locusta mig
13	16	25.0	9	1	DNF1_LOCMI	P16339 locusta mig
14	16	25.0	10	1	MOSQ_CLYJA	P19962 clypeaster
15	16	25.0	12	1	CALM_TETTH	Q05055 tetrahymena
16	15	23.4	7	1	UH11_RAT	P56576 rattus norv
17	15	23.4	8	1	CPD1_ENTFA	P13269 enterococcu
18	15	23.4	8	1	HTF_TENMO	P25419 tenebrio mo
19	15	23.4	8	1	RPCH_PANBO	P08939 pandalus bo
20	15	23.4	8	1	UPAA_HUMAN	P30096 homo sapien
21	15	23.4	9	1	XYLA_STRSQ	P19149 streptomyce
22	15	23.4	13	1	CRBL_VESMA	P17232 vespa manda
23	14	21.9	8	1	ALL4_CYDPO	P82155 cydia pomon
24	14	21.9	9	1	CONO_CONST	P05487 conus stria
25	14	21.9	9	1	HUTU_KLEAE	P12381 klebsiella
26	14	21.9	10	1	ESTA_SCHGA	P81012 schizaphis
27	14	21.9	10	1	TKL2_LOCMI	P16224 locusta mig
28	14	21.9	10	1	URE3_MORMO	P17339 morganella
29	14	21.9	11	1	MORN_HUMAN	P01163 homo sapien
30	14	21.9	12	1	FRE1_LITIN	P82021 litoria inf
31	14	21.9	13	1	CP1_APLCA	Q10998 aplysia cal
32	14	21.9	13	1	HPA1_RANES	P32415 rana escule
33	14	21.9	13	1	NP1_LYMST	P80178 lymnaea sta
34	14	21.9	13	1	NP2_LYMST	P80179 lymnaea sta
35	14	21.9	13	1	NP4_LYMST	P80181 lymnaea sta
36	14	21.9	13	1	NP5_LYMST	P80182 lymnaea sta
37	14	21.9	13	1	SA2A_ONCMY	P82238 oncorhynchu
38	14	21.9	13	1	SA2B_ONCMY	P82239 oncorhynchu
39	13	20.3	9	1	CCAP_CARMA	P38556 carcinus ma
40	13	20.3	9	1	FAR9_ASCSU	P43172 ascaris suu
41	13	20.3	9	1	LMT3_LOCMI	P41489 locusta mig
42	13	20.3	9	1	OXYT_BUFRE	P42995 bufo regula
43	13	20.3	9	1	OXYT_EISFO	P42998 eisenia foe
44	13	20.3	10	1	TKU1_UREUN	P40751 urechis uni
45	13	20.3	13	1	CH60_CANFA	P49818 canis famil

ALIGNMENTS

RESULT 1
ES1_RAT
ID ES1_RAT STANDARD; PRT; 11 AA.
AC P56571;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ES1 protein, mitochondrial (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RC STRAIN=WISTAR; TISSUE=Heart;
RA Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA Jungblut P.R.;
RL Submitted (SEP-1998) to the SWISS-PROT data bank.
CC -!- SUBCELLULAR LOCATION: Mitochondrial (Potential).
CC -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN
CC (SPOT P2) IS: 8.9, ITS MW IS: 25 kDa.
CC -!- SIMILARITY: BELONGS TO THE ES1 FAMILY.
KW Mitochondrion.
FT NON_TER 11 11
SQ SEQUENCE 11 AA; 1142 MW; D862272D32C72DC2 CRC64;

Query Match 35.9%; Score 23; DB 1; Length 11;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 LVLSGS 7
          |||||:
Db      6 LVLSGT 11
```

RESULT 2
AKH_TABAT
ID AKH_TABAT STANDARD; PRT; 8 AA.
AC P14595;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Adipokinetic hormone (AKH) (Dipteran corpora cardiaca factor I)
DE (DCC I).
OS Tabanus atratus (Horse fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;
OC Tabanidae; Tabanus.
OX NCBI_TaxID=7207;
RN [1]
RP SEQUENCE.
RC TISSUE=Corpora cardiaca;
RX MEDLINE=90046758; PubMed=2813385;
RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,
RA Vogel V.W., Zhang Y.-S., Hayes D.K.;

```
RT   "Primary structure of two neuropeptide hormones with adipokinetic and
RT   hypotrehalosemic activity isolated from the corpora cardiaca of horse
RT   flies (Diptera).";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A33995; A33995.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  949 MW;  86786771A9D1A736 CRC64;

  Query Match             31.2%;  Score 20;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   3
HTF2_PERAM
ID   HTF2_PERAM     STANDARD;      PRT;     8 AA.
AC   P04549;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic factor II (Neuropeptide M-II) (Periplanetin CC-2)
DE   (PeA-CAH-II) (LeD-CC-II) (Hypertrehalosaemic neuropeptide II).
OS   Periplaneta americana (American cockroach),
OS   Leptinotarsa decemlineata (Colorado potato beetle), and
OS   Blatta orientalis (Oriental cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
OX   NCBI_TaxID=6978, 7539, 6976;
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.americana;
RX   MEDLINE=85046530; PubMed=6548628;
RA   Witten J.L., Schaffer M.H., O'Shea M., Cook J.C., Hemling M.E.,
RA   Rinehart K.L. Jr.;
RT   "Structures of two cockroach neuropeptides assigned by fast atom
RT   bombardment mass spectrometry.";
RL   Biochem. Biophys. Res. Commun. 124:350-358(1984).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.americana;
RX   MEDLINE=84298179; PubMed=6591205;
RA   Scarborough R.M., Jamieson G.C., Kalish F., Kramer S.J., McEnroe G.A.,
RA   Miller C.A., Schooley D.A.;
RT   "Isolation and primary structure of two peptides with
RT   cardioacceleratory and hyperglycemic activity from the corpora
RT   cardiaca of Periplaneta americana.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:5575-5579(1984).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.decemlineata; TISSUE=Corpora cardiaca;
RX   MEDLINE=90160053; PubMed=2576128;
RA   Gaede G., Kellner R.;
RT   "The metabolic neuropeptides of the corpus cardiacum from the potato
RT   beetle and the American cockroach are identical.";
RL   Peptides 10:1287-1289(1989).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=B.orientalis; TISSUE=Corpora cardiaca;
RX   MEDLINE=90253659; PubMed=2340112;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structures of hypertrehalosaemic neuropeptides isolated from
RT   the corpora cardiaca of the cockroaches Leucophaea maderae,
RT   Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT   and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT   atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 371:345-354(1990).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A05170; A05170.
DR   PIR; S08996; S08996.
DR   PIR; B44960; B44960.
DR   PIR; B49823; B49823.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1006 MW;  86745771A9D1A736 CRC64;

  Query Match             31.2%;  Score 20;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   4
HTF2_CARMO
ID   HTF2_CARMO     STANDARD;      PRT;    10 AA.
AC   P11385;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hypertrehalosaemic factor II (HTF-II) (HRTH-II) (Hypertrehalosaemic
DE   neuropeptide II).
OS   Carausius morosus (Indian stick insect), and
OS   Extatosoma tiaratum (Stick insect).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Phasmatodea; Heteronemiidae;
OC   Carausius.
OX   NCBI_TaxID=7022, 7024;
RN   [1]
RP   SEQUENCE.
RC   SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX   MEDLINE=87157103; PubMed=3828078;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structure of the hypertrehalosaemic factor II from the
RT   corpus cardiacum of the Indian stick insect, Carausius morosus,
RT   determined by fast atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 368:67-75(1987).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=E.tiaratum; TISSUE=Corpora cardiaca;
RX   MEDLINE=90253659; PubMed=2340112;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structures of hypertrehalosaemic neuropeptides isolated from
RT   the corpora cardiaca of the cockroaches Leucophaea maderae,
RT   Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT   and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT   atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 371:345-354(1990).
RN   [3]
RP   CARBOHYDRATE-LINKAGE SITE.
RC   SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX   MEDLINE=93129188; PubMed=1482345;
RA   Gaede G., Kellner R., Rinehart K.L. Jr., Proefke M.L.;
```

```
RT   "A tryptophan-substituted member of the AKH/RPCH family isolated from
RT   a stick insect corpus cardiacum.";
RL   Biochem. Biophys. Res. Commun. 189:1303-1309(1992).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- MASS SPECTROMETRY: MW=1308.61; METHOD=FAB.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S07157; S07157.
DR   PIR; S09138; S09138.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Glycoprotein.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   CARBOHYD      8      8       C-LINKED (MAN) (PROBABLE).
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1164 MW;  9B9036745771A9D1 CRC64;


  Query Match             31.2%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   5
HTF_HELZE
ID   HTF_HELZE      STANDARD;      PRT;     10 AA.
AC   P16353;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic hormone (HeZ-HRTH).
OS   Heliothis zea (Corn earworm) (Bollworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
OX   NCBI_TaxID=7113;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=88326324; PubMed=3415690;
RA   Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Bird T.G.,
RA   Tseng C.M., Zhang Y.S., Hayes D.K.;
RT   "Isolation and primary structure of a neuropeptide hormone from
RT   Heliothis zea with hypertrehalosemic and adipokinetic activities.";
RL   Biochem. Biophys. Res. Commun. 155:344-350(1988).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A31571; A31571.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1096 MW;  8E70367865A5B9D1 CRC64;


  Query Match             31.2%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   6
HTF_TABAT
ID   HTF_TABAT      STANDARD;      PRT;     10 AA.
AC   P14596;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic factor (HOTH) (Dipteran corpora cardiaca factor II)
DE   (DCC II).
OS   Tabanus atratus (Horse fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;
OC   Tabanidae; Tabanus.
OX   NCBI_TaxID=7207;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=90046758; PubMed=2813385;
RA   Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,
RA   Vogel V.W., Zhang Y.-S., Hayes D.K.;
RT   "Primary structure of two neuropeptide hormones with adipokinetic and
RT   hypotrehalosemic activity isolated from the corpora cardiaca of horse
RT   flies (Diptera).";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; B33995; B33995.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1169 MW;  916036786771A9D1 CRC64;


  Query Match             31.2%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 QLTF 13
           ||||
Db       1 QLTF 4


RESULT   7
NP3_LYMST
ID   NP3_LYMST      STANDARD;      PRT;     13 AA.
AC   P80180;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-JUL-1993 (Rel. 26, Last annotation update)
DE   Lymna-DF-amide 3.
OS   Lymnaea stagnalis (Great pond snail).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC   Lymnaeidae; Lymnaea.
OX   NCBI_TaxID=6523;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Ganglion;
RX   MEDLINE=93238777; PubMed=8477756;
RA   Johnsen A.H., Rehfeld J.F.;
RT   "LymnaDFamides, a new family of neuropeptides from the pond snail,
RT   Lymnaea stagnalis. Clue to cholecystokinin immunoreactivity in
RT   invertebrates?";
RL   Eur. J. Biochem. 213:875-879(1993).
CC   -!- SIMILARITY: RELATED TO THE CHOLECYSTOKININ (CCK) FAMILY.
DR   PIR; S32473; S32473.
KW   Neuropeptide; Amidation.
FT   MOD_RES      13     13       AMIDATION.
FT   UNSURE       12     12
SQ   SEQUENCE   13 AA;  1462 MW;  9CA07BA3F5D5B865 CRC64;
```

```
 Query Match             31.2%;  Score 20;  DB 1;  Length 13;
 Best Local Similarity   80.0%;  Pred. No. 7.2e+02;
 Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 LSGSA 8
           :||||
Db       5 ISGSA 9


RESULT   8
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;      8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   Sex pheromone CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85051889; PubMed=6437872;
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;


  Query Match             29.7%;  Score 19;  DB 1;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LVLSG 6
           |||:|
Db       4 LVLAG 8


RESULT   9
ITB5_BOVIN
ID   ITB5_BOVIN     STANDARD;      PRT;     13 AA.
AC   P80747;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Integrin beta-5 (Fragment).
GN   ITGB5.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Mammary gland;
RX   MEDLINE=97299777; PubMed=9154926;
RA   Andersen M.H., Berglund L., Rasmussen J.T., Petersen T.E.;
RT   "Bovine PAS-6/7 binds alpha v beta 5 integrins and anionic
RT   phospholipids through two domains.";
RL   Biochemistry 36:5441-5446(1997).
CC   -!- FUNCTION: INTEGRIN ALPHA-V/BETA-5 IS A RECEPTOR FOR FIBRONECTIN.
CC       IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGAND.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-5
CC       ASSOCIATES WITH ALPHA-V.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
DR   InterPro; IPR001169; Integrin_beta_C.
DR   PROSITE; PS00243; INTEGRIN_BETA; PARTIAL.
KW   Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW   Repeat.
FT   NON_TER      13      13
SQ   SEQUENCE   13 AA;  1299 MW;  844197D005B9B865 CRC64;


  Query Match             28.1%;  Score 18;  DB 1;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 SGSA 8
           ||||
Db       6 SGSA 9


RESULT  10
TEMA_RANTE
ID   TEMA_RANTE     STANDARD;      PRT;     13 AA.
AC   P56917;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Temporin A.
OS   Rana temporaria (European common frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8407;
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=Skin;
RX   MEDLINE=97175050; PubMed=9022710;
RA   Simmaco M., Mignogna G., Canofeni S., Miele R., Mangoni M.L.,
RA   Barra D.;
RT   "Temporins, antimicrobial peptides from the European red frog Rana
RT   temporaria.";
RL   Eur. J. Biochem. 242:788-792(1996).
CC   -!- FUNCTION: HAS ANTIBACTERIAL ACTIVITY AGAINST GRAM-POSITIVE
CC       BACTERIA.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAEGURIN/RUGOSIN
CC       FAMILY.
KW   Amphibian skin; Antibiotic; Amidation; Multigene family.
FT   MOD_RES      13      13       AMIDATION.
SQ   SEQUENCE   13 AA;  1398 MW;  2653612B9DECD408 CRC64;


  Query Match             28.1%;  Score 18;  DB 1;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VLSG 6
           ||||
Db       8 VLSG 11


RESULT  11
TEMF_RANTE
ID   TEMF_RANTE     STANDARD;      PRT;     13 AA.
AC   P56921;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Temporin F.
OS   Rana temporaria (European common frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8407;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Skin;
RX   MEDLINE=97175050; PubMed=9022710;
```

```
RA   Simmaco M., Mignogna G., Canofeni S., Miele R., Mangoni M.L.,
RA   Barra D.;
RT   "Temporins, antimicrobial peptides from the European red frog Rana
RT   temporaria.";
RL   Eur. J. Biochem. 242:788-792(1996).
CC   -!- FUNCTION: HAS ANTIBACTERIAL ACTIVITY AGAINST GRAM-NEGATIVE AND
CC       GRAM-POSITIVE BACTERIA.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAEGURIN/RUGOSIN
CC       FAMILY.
KW   Amphibian skin; Antibiotic; Amidation; Multigene family.
FT   MOD_RES      13      13       AMIDATION.
SQ   SEQUENCE   13 AA;  1370 MW;  2653612B9DECC338 CRC64;


  Query Match             28.1%;  Score 18;  DB 1;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VLSG 6
          ||||
Db      8 VLSG 11


RESULT  12
AKHX_LOCMI
ID   AKHX_LOCMI     STANDARD;      PRT;     10 AA.
AC   P81626;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Peptide hormone.
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RA   Siegert K.J.;
RL   Submitted (DEC-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE REGULATION OF LOCUST
CC       INTERMEDIARY METABOLISM, BEHAVIOR AND/OR DEVELOPMENT.
CC   -!- SIMILARITY: SOME SIMILARITY TO THE AKH / HRTH / RPCH FAMILY.
DR   InterPro; IPR002047; AKH.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1       1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10      10       AMIDATION.
SQ   SEQUENCE   10 AA;  1222 MW;  81BFF67AB415B9D1 CRC64;


  Query Match             26.6%;  Score 17;  DB 1;  Length 10;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+03;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 QLTF 13
          |:||
Db      1 QVTF 4


RESULT  13
DNF1_LOCMI
ID   DNF1_LOCMI     STANDARD;      PRT;      9 AA.
AC   P16339;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Locupressin (Diuretic neuropeptide F1/F2).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Suboesophageal ganglion, and Thoracic ganglion;
RX   MEDLINE=88077077; PubMed=3689410;
RA   Proux J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,
RA   Delaage M., Schooley D.A.;
RT   "Identification of an arginine vasopressin-like diuretic hormone from
RT   Locusta migratoria.";
RL   Biochem. Biophys. Res. Commun. 149:180-186(1987).
CC   -!- FUNCTION: DIURETIC HORMONE.
CC   -!- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF F1.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A29477; A29477.
DR   InterPro; IPR000981; Neurhypophys_horm.
DR   Pfam; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Neuropeptide; Amidation.
FT   DISULFID      1       6       IN F1.
FT   DISULFID      1       1       INTERCHAIN (WITH C-6') (IN F2).
FT   DISULFID      6       6       INTERCHAIN (WITH C-1') (IN F2).
FT   MOD_RES       9       9       AMIDATION.
SQ   SEQUENCE   9 AA;  976 MW;  56EB176EB451A057 CRC64;


  Query Match             25.0%;  Score 16;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CLV 3
          ||:
Db      1 CLI 3


RESULT  14
MOSQ_CLYJA
ID   MOSQ_CLYJA     STANDARD;      PRT;     10 AA.
AC   P19962;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [Gln-6]-mosact.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
OX   NCBI_TaxID=7644;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Egg jelly;
RA   Suzuki N., Kurita M., Yoshino K., Kajiura H., Nomura K., Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0025; JN0025.
SQ   SEQUENCE   10 AA;  1019 MW;  9AFB032456DDC5BA CRC64;


  Query Match             25.0%;  Score 16;  DB 1;  Length 10;
  Best Local Similarity   57.1%;  Pred. No. 3.3e+03;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 SGSARQL 11
          | ||: |
Db      2 SDSAQNL 8


RESULT  15
CALM_TETTH
ID   CALM_TETTH     STANDARD;      PRT;     12 AA.
AC   Q05055;
DT   01-FEB-1994 (Rel. 28, Created)
```

```
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Calmodulin (Fragment).
OS   Tetrahymena thermophila.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281388; PubMed=8506136;
RA   Katoh M., Hirono M., Takemasa T., Kimura M., Watanabe Y.;
RT   "A micronucleus-specific sequence exists in the 5'-upstream region of
RT   calmodulin gene in Tetrahymena thermophila.";
RL   Nucleic Acids Res. 21:2409-2414(1993).
CC   -!- FUNCTION: CALMODULIN MEDIATES THE CONTROL OF A LARGE NUMBER OF
CC       ENZYMES BY CA(++). AMONG THE ENZYMES TO BE STIMULATED BY THE
CC       CALMODULIN-CA(++) COMPLEX ARE A NUMBER OF PROTEIN KINASES AND
CC       PHOSPHATASES.
CC   -!- MISCELLANEOUS: THIS PROTEIN HAS FOUR FUNCTIONAL CALCIUM-BINDING
CC       SITES.
CC   -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12774; BAA02239.1; -.
DR   HSSP; P02593; 2CLN.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; PARTIAL.
KW   Calcium-binding; Repeat; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1393 MW;  83F31CD443DB1B01 CRC64;


  Query Match             25.0%;  Score 16;  DB 1;  Length 12;
  Best Local Similarity   80.0%;  Pred. No. 4e+03;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 ARQLT 12
           | |||
Db       1 ADQLT 5


Search completed: June  6, 2002, 06:26:36
Job time: 189 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:07 ; Search time 23.81 Seconds
(without alignments)
94.453 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 2182

Minimum DB seq length: 0
Maximum DB seq length: 13

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	32.8	10	5	Q25356	Q25356 locusta mig
2	21	32.8	11	11	Q9R0K9	Q9r0k9 mus musculu
3	21	32.8	12	10	Q945C3	Q945c3 crypthecodi
4	19	29.7	13	4	Q9UM84	Q9um84 homo sapien
5	19	29.7	13	5	O61340	O61340 panulirus i
6	18	28.1	8	2	O09258	O09258 synechococc
7	18	28.1	10	4	Q9P2Z9	Q9p2z9 homo sapien
8	18	28.1	11	11	P97330	P97330 mus musculu
9	18	28.1	12	11	Q64296	Q64296 mus musculu
10	18	28.1	13	6	Q9TQS1	Q9tqs1 equus cabal
11	18	28.1	13	13	Q9PWP4	Q9pwp4 dissostichu
12	17	26.6	9	13	Q92009	Q92009 gallus gall
13	17	26.6	11	15	Q98YS3	Q98ys3 human immun
14	17	26.6	12	2	P83054	P83054 bacteroides
15	17	26.6	12	6	Q9TQY4	Q9tqy4 bos taurus
16	17	26.6	13	6	Q9TUD7	Q9tud7 bos taurus
17	17	26.6	13	10	Q9XIW6	Q9xiw6 oryza sativ
18	17	26.6	13	11	Q62352	Q62352 mus musculu
19	17	26.6	13	11	Q62354	Q62354 mus musculu
20	17	26.6	13	11	Q62355	Q62355 mus musculu
21	16	25.0	8	4	Q96RN9	Q96rn9 homo sapien
22	16	25.0	8	12	Q89498	Q89498 murine hepa
23	16	25.0	8	12	Q83349	Q83349 murine coro
24	16	25.0	9	11	O35953	O35953 mus musculu
25	16	25.0	9	12	Q9PYK1	Q9pyk1 simian viru
26	16	25.0	9	12	Q9IBM8	Q9ibm8 simian viru
27	16	25.0	10	2	Q9X3M2	Q9x3m2 prochloroco
28	16	25.0	10	8	Q37103	Q37103 saccharomyc
29	16	25.0	11	5	P82700	P82700 leucophaea
30	16	25.0	12	10	Q945C2	Q945c2 crypthecodi
31	16	25.0	12	12	O10421	O10421 influenza a
32	16	25.0	13	8	Q94RE2	Q94re2 leptomonas
33	16	25.0	13	10	Q9S8N1	Q9s8n1 hordeum vul
34	16	25.0	13	11	P97944	P97944 mus musculu
35	16	25.0	13	11	Q9QVD5	Q9qvd5 rattus sp.
36	16	25.0	13	12	Q9WMG5	Q9wmg5 sigma virus
37	15	23.4	7	2	P70804	P70804 azotobacter
38	15	23.4	8	6	P82929	P82929 bos taurus
39	15	23.4	8	11	Q60615	Q60615 mus musculu
40	15	23.4	8	12	O89965	O89965 polyomaviru
41	15	23.4	8	13	P82079	P82079 limnodynast
42	15	23.4	9	5	Q27396	Q27396 babesia bov
43	15	23.4	9	6	Q9TRS0	Q9trs0 oryctolagus
44	15	23.4	10	10	Q08938	Q08938 nicotiana t
45	15	23.4	10	10	Q94II9	Q94ii9 zea mays (m

ALIGNMENTS

```
RESULT   1
Q25356
ID   Q25356       PRELIMINARY;       PRT;    10 AA.
AC   Q25356;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   VITELLOGENIN B (FRAGMENT).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88003971; PubMed=2820677;
RA   Locke J., White B.N., Wyatt G.R.;
RT   "Cloning and 5' end nucleotide sequences of two juvenile hormone-
RT   inducible vitellogenin genes of the African migratory locust.";
RL   DNA 6:331-342(1987).
DR   EMBL; M17334; AAA29285.1; -.
FT   NON_TER      10      10
SQ   SEQUENCE   10 AA;  1116 MW;  D1B31165B720572D CRC64;


  Query Match             32.8%;  Score 21;  DB 5;  Length 10;
  Best Local Similarity   80.0%;  Pred. No. 1.8e+03;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LVLSG 6
            |:|||
Db       4 LILSG 8


RESULT   2
Q9R0K9
ID   Q9R0K9       PRELIMINARY;       PRT;    11 AA.
AC   Q9R0K9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BCL-2 ASSOCIATED X PROTEIN (FRAGMENT).
GN   BAX OR BAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20035745; PubMed=10570968;
RA   Igata E., Inoue T., Ohtani-Fujita N., Sowa Y., Tsujimoto Y., Sakai T.;
RT   "Molecular cloning and functional analysis of the murine bax gene
RT   promoter.";
RL   Gene 238:407-415(1999).
DR   EMBL; AB029557; BAA82406.1; -.
DR   MGD; MGI:99702; Bax.
FT   NON_TER      11     11
SQ   SEQUENCE   11 AA;  1037 MW;  90AAB34E36DB1865 CRC64;


  Query Match             32.8%;  Score 21;  DB 11;  Length 11;
  Best Local Similarity   50.0%;  Pred. No. 2e+03;
  Matches    4;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 LSGSARQL 11
            : ||  ||
Db       1 MDGSGEQL 8


RESULT   3
Q945C3
ID   Q945C3      PRELIMINARY;      PRT;    12 AA.
AC   Q945C3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   P80 PROTEIN (FRAGMENT).
OS   Crypthecodinium cohnii (Dinoflagellate).
OC   Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Crypthecodiniaceae;
OC   Crypthecodinium.
OX   NCBI_TaxID=2866;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99433238; PubMed=10505419;
RA   Ausseil J., Soyer-Gobillard M.O., Geraud M.L., Bhaud Y., Baines I.,
RA   Preston T., Moreau H.;
RT   "Characterization of p80, a novel nuclear and cytoplasmic protein in
RT   dinoflagellates.";
RL   Protist 150:197-211(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Guillebault D., Derelle E., Lozano J.C., Bingham S., Moreau H.;
RT   "A single TBP-like protein is present in the marine unicellular
RT   organism: the dinoflagellate Crypthecodinium cohnii.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF417568; AAL15906.1; -.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1364 MW;  615BF873FE204414 CRC64;


  Query Match             32.8%;  Score 21;  DB 10;  Length 12;
  Best Local Similarity   57.1%;  Pred. No. 2.2e+03;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 SARQLTF 13
            |||:: |
Db       3 SARRILF 9


RESULT   4
Q9UM84
ID   Q9UM84      PRELIMINARY;      PRT;    13 AA.
AC   Q9UM84;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RET-ELE1 PROTEIN (FRAGMENT).
GN   RET-ELE1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96400323; PubMed=8806699;
RA   Fugazzola L., Pierotti M.A., Vigano E., Pacini F., Vorontsova T.V.,
RA   Bongarzone I.;
RT   "Molecular and biochemical analysis of RET/PTC4, a novel oncogenic
RT   rearrangement between RET and ELE1 genes, in a post-Chernobyl
RT   papillary thyroid cancer.";
RL   Oncogene 13:1093-1097(1996).
DR   EMBL; S83049; AAD14423.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   13 AA;  1463 MW;  7C84ED9161C1A69B CRC64;


  Query Match             29.7%;  Score 19;  DB 4;  Length 13;
  Best Local Similarity   42.9%;  Pred. No. 5.8e+03;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 SARQLTF 13
            |: ::||
Db       3 SSAEMTF 9


RESULT   5
O61340
ID   O61340      PRELIMINARY;      PRT;    13 AA.
AC   O61340;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POTASSIUM CHANNEL (FRAGMENT).
GN   SHAKER.
OS   Panulirus interruptus (California spiny lobster).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC   Palinuroidea; Palinuridae; Panulirus.
OX   NCBI_TaxID=6735;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98330950; PubMed=9666521;
RA   Kim M., Baro D.J., Lanning C.C., Doshi M., Moskowitz H.S., Farnham J.,
RA   Harris-Warrick R.M.;
RT   "Expression of Panulirus shaker potassium channel splice variants.";
RL   Recept. Channels 5:291-304(1998).
DR   EMBL; AF017135; AAC05915.1; -.
FT   NON_TER      13     13
SQ   SEQUENCE   13 AA;  1336 MW;  CBA864F1E31E31AD CRC64;


  Query Match             29.7%;  Score 19;  DB 5;  Length 13;
  Best Local Similarity   80.0%;  Pred. No. 5.8e+03;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LVLSG 6
            |:|||
Db       8 LLLSG 12


RESULT   6
O09258
ID   O09258      PRELIMINARY;      PRT;     8 AA.
AC   O09258;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
```

```
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NIFH (FRAGMENT).
GN   NIFH.
OS   Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothece PCC 8801).
OC   Bacteria; Cyanobacteria; Chroococcales; Cyanothece.
OX   NCBI_TaxID=41431;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RF-1;
RX   MEDLINE=99231861; PubMed=10217509;
RA   Huang T.C., Lin R.F., Chu M.K., Chen H.M.;
RT   "Organization and expression of nitrogen-fixation genes in the aerobic
RT   nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain
RT   RF-1.";
RL   Microbiology 145:743-753(1999).
DR   EMBL; AF001780; AAC33369.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  985 MW;  F16B59CDD046C406 CRC64;

  Query Match             28.1%;  Score 18;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 5.6e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 RQLTF 13
           ||: |
Db       2 RQIAF 6


RESULT   7
Q9P2Z9
ID   Q9P2Z9      PRELIMINARY;      PRT;    10 AA.
AC   Q9P2Z9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EPIDERMAL GROWTH FACTOR RECEPTOR (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91337091; PubMed=1678600;
RA   Humphrey P.A., Gangarosa L.M., Wong A.J., Archer G.E.,
RA   Lund-Johansen M., Bjerkvig R., Laerum O.D., Friedman H.S.,
RA   Bigner D.D.;
RT   "Deletion-mutant epidermal growth factor receptor in human gliomas:
RT   effects of type II mutation on receptor function.";
RL   Biochem. Biophys. Res. Commun. 178:1413-1420(1991).
DR   EMBL; S51343; AAB19486.2; -.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER      10     10
SQ   SEQUENCE   10 AA;  1020 MW;  439DEEAEB87B1727 CRC64;

  Query Match             28.1%;  Score 18;  DB 4;  Length 10;
  Best Local Similarity   50.0%;  Pred. No. 6.8e+03;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           | :| |
Db       1 CKLLEG 6


RESULT   8
P97330
ID   P97330      PRELIMINARY;      PRT;    11 AA.
AC   P97330;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CD45-AP (LSM-1).
GN   PTPRCAP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97124850; PubMed=8954783;
RA   Bruyns E., Mincheva A., Bruyns R.M., Kirchgessner H., Weitz S.,
RA   Lichter P., Meuer S., Schraven B.;
RT   "Sequence, genomic organization and chromosomal Localization of the
RT   human LPAP (PTPRCAP) and Mouse CD45-AP/LSM-1 genes.";
RL   Genomics 38:79-83(1996).
DR   EMBL; X97268; CAA65923.1; -.
DR   MGD; MGI:97811; Ptprcap.
SQ   SEQUENCE   11 AA;  1150 MW;  50695413B5A772C7 CRC64;

  Query Match             28.1%;  Score 18;  DB 11;  Length 11;
  Best Local Similarity   80.0%;  Pred. No. 7.5e+03;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLS 5
           ||| |
Db       3 CLVPS 7


RESULT   9
Q64296
ID   Q64296      PRELIMINARY;      PRT;    12 AA.
AC   Q64296;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CD44 ANTIGEN (FRAGMENT).
GN   CD44.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RA   Yu Q., Toole B.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RA   Yu Q., Toole B.P.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U57612; AAC52806.1; -.
DR   EMBL; U57611; AAB08756.1; -.
DR   MGD; MGI:88338; Cd44.
FT   NON_TER       1      1
SQ   SEQUENCE   12 AA;  1235 MW;  CBFDEB7444ADC2D2 CRC64;

  Query Match             28.1%;  Score 18;  DB 11;  Length 12;
  Best Local Similarity   75.0%;  Pred. No. 8.2e+03;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVL 4
           |||:
Db       2 CLVV 5


RESULT  10
Q9TQS1
ID   Q9TQS1      PRELIMINARY;      PRT;    13 AA.
AC   Q9TQS1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TRANSFERRIN (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Giffard J.M., Brandon R.B., Bell T.K.;
RT   "Further identification of single nucleotide polymorphisms in the
RT   equine transferrin gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF185747; AAF05461.1; -.
DR   EMBL; AF185746; AAF05460.1; -.
FT   NON_TER       1      1
FT   NON_TER      13     13
SQ   SEQUENCE   13 AA;  1351 MW;  D893DAE970B6CEBD CRC64;


  Query Match             28.1%;  Score 18;  DB 6;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 SGSA 8
           ||||
Db       3 SGSA 6


RESULT  11
Q9PWP4
ID   Q9PWP4       PRELIMINARY;       PRT;    13 AA.
AC   Q9PWP4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CHIMERIC AFGP/TRYPSINOGEN-LIKE SERINE PROTEASE (FRAGMENT).
OS   Dissostichus mawsoni (Antartic cod).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC   Notothenioidei; Nototheniidae; Dissostichus.
OX   NCBI_TaxID=36200;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99447034; PubMed=10519545;
RA   Cheng C.H., Chen L.;
RT   "Evolution of an antifreeze glycoprotein.";
RL   Nature 401:443-444(1999).
DR   EMBL; AF134320; AAD37246.1; -.
KW   Protease.
FT   NON_TER      13     13
SQ   SEQUENCE   13 AA;  1340 MW;  5186FA54AF1E2727 CRC64;


  Query Match             28.1%;  Score 18;  DB 13;  Length 13;
  Best Local Similarity   57.1%;  Pred. No. 9e+03;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LVLSGSA 8
           |:| |:|
Db       7 LLLIGAA 13


RESULT  12
Q92009
ID   Q92009       PRELIMINARY;       PRT;     9 AA.
AC   Q92009;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   C-REL PROTEIN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91133738; PubMed=2284104;
RA   Hannink M., Temin H.M.;
RT   "Structure and autoregulation of the c-rel promoter.";
RL   Oncogene 5:1843-1850(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hannink M., Temin H.M.;
RL   Oncogene 0:0-0(1990).
DR   EMBL; X56440; CAA39822.1; -.
DR   EMBL; X56515; CAA39866.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE    9 AA;  805 MW;  DE317DD87865A2CD CRC64;


  Query Match             26.6%;  Score 17;  DB 13;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 5.6e+05;
  Matches    3;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LVLSGSA 8
           : :|| |
Db       1 MAVSGGA 7


RESULT  13
Q98YS3
ID   Q98YS3       PRELIMINARY;       PRT;    11 AA.
AC   Q98YS3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TRUNCATED POL PROTEIN (FRAGMENT).
GN   POL.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=985829;
RA   Schmidt B., Walter H., Moschik G., Paatz C., Werwein M., Schwingel E.,
RA   Korn K.;
RT   "Recovery of HIV-1 pol gene sequences by direct sequencing of
RT   amplification products derived from plasma samples.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF347394; AAK32471.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   11 AA;  1195 MW;  E96941B8D878773A CRC64;


  Query Match             26.6%;  Score 17;  DB 15;  Length 11;
  Best Local Similarity   33.3%;  Pred. No. 1.2e+04;
  Matches    3;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVLSGSARQ 10
           : | |  |:
Db       3 IXLGGQLRE 11


RESULT  14
P83054
ID   P83054       PRELIMINARY;       PRT;    12 AA.
AC   P83054;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HEPARIN LYASE (EC 4.2.2.7) (HEPARIN ELIMINASE) (HEPARINASE)
DE   (FRAGMENT).
OS   Bacteroides stercoris.
```

```
OC   Bacteria; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC   Bacteroides.
OX   NCBI_TaxID=46506;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=HJ-15;
RX   MEDLINE=20381043; PubMed=10920269;
RA   Kim B.-T., Kim W.-S., Kim Y.S., Linhardt R.J., Kim D.-H.;
RT   "Purification and characterization of a novel heparinase from
RT   Bacteroides stercoris HJ-15.";
RL   J. Biochem. 128:323-328(2000).
CC   -!- FUNCTION: DEGRADES HEPARIN AND HEPARAN SULFATE.
CC   -!- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF POLYSACCHARIDES
CC       CONTAINING 1,4-LINKED GLUCURONATE OR IDURONATE RESIDUES AND 1,4-
CC       ALPHA-LINKED 2-SULFOAMINO-2-DEOXY-6-SULFO-D-GLUCOSE RESIDUES TO
CC       GIVE OLIGOSACCHARIDES WITH TERMINAL 4-DEOXY-ALPHA-D-GLUC-4-
CC       ENURONOSYL GROUPS AT THEIR NON-REDUCING ENDS.
CC   -!- ENZYME REGULATION: INHIBITED BY CUPRIC ION, LEAD AND SOME AGENTS
CC       THAT MODIFY HISTIDINE AND CYSTEINE RESIDUES. ACTIVATED BY KCL AND
CC       BY REDUCING AGENTS, SUCH AS DITHIOTHREITOL AND 2-MERCAPTOETHANOL.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- MISCELLANEOUS: THE ENZYME HAS AN ISOELECTRIC POINT OF 8.7. ITS
CC       OPTIMUM PH IS 7.2 AND THE OPTIMUM TEMPERATURE IS 45 DEGREES
CC       CELSIUS.
KW   Lyase; Heparin-binding.
FT   NON_TER       1      1
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1381 MW;  CD9CCD8B98F6D72D CRC64;

  Query Match             26.6%;  Score 17;  DB 2;  Length 12;
  Best Local Similarity   40.0%;  Pred. No. 1.3e+04;
  Matches    4;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 LSGSARQLTF 13
           ::  | | ||
Db       1 MADEALQHTF 10


RESULT  15
Q9TQY4
ID   Q9TQY4      PRELIMINARY;      PRT;    12 AA.
AC   Q9TQY4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GLYCOPROTEIN H-B N-TERMINAL, GPH-B N-TERMINAL=KEX2/SUBTILISIN-RELATED
DE   PROTEASE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91340701; PubMed=1874725;
RA   Christie D.L., Batchelor D.C., Palmer D.J.;
RT   "Identification of kex2-related proteases in chromaffin granules by
RT   partial amino acid sequence analysis.";
RL   J. Biol. Chem. 266:15679-15683(1991).
SQ   SEQUENCE   12 AA;  1303 MW;  9F2FF9E2782DC5BA CRC64;

  Query Match             26.6%;  Score 17;  DB 6;  Length 12;
  Best Local Similarity   55.6%;  Pred. No. 1.3e+04;
  Matches    5;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 VLSGSARQL 11
           ||  ||  |
Db       2 VLXDSALDL 10
```

Search completed: June 6, 2002, 06:26:20
Job time: 193 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:26 ; Search time 29.12 Seconds
(without alignments)
49.587 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 166670

Minimum DB seq length: 0
Maximum DB seq length: 13

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	64	100.0	13	19	AAW47590	T-cell receptor CD
2	42.5	66.4	12	19	AAW47591	T-cell receptor CD
3	26.5	41.4	10	22	AAG86220	Saccharomyces cere
4	26.5	41.4	10	22	AAG86310	Saccharomyces cere
5	26	40.6	9	22	AAB46917	Human SART-1 deriv
6	26	40.6	10	22	AAG84200	Arabidopsis thalia
7	26	40.6	10	22	AAG84202	Arabidopsis thalia
8	26	40.6	11	19	AAW54287	Cytohesin 1 PH dom
9	25	39.1	9	20	AAY48993	Membrane dipeptida
10	25	39.1	9	22	AAB45862	Human tumor-associ
11	25	39.1	10	22	AAG96480	Human complementar
12	25	39.1	10	22	AAG96482	Human complementar
13	25	39.1	12	18	AAW32678	Human platelet gly
14	25	39.1	12	19	AAW71808	Mimotope capable o
15	25	39.1	12	22	AAB84372	Splice-variance re
16	24	37.5	7	22	AAB80722	Human glandular ka
17	24	37.5	8	22	AAM22237	HIV peptide SEQ ID
18	24	37.5	10	22	AAG94861	Human complementar
19	24	37.5	10	22	AAG94863	Human complementar
20	24	37.5	10	22	AAG95948	Human complementar
21	24	37.5	10	22	AAG96400	Human complementar
22	24	37.5	12	15	AAR60702	N-terminal a.a. of
23	24	37.5	12	21	AAY96716	Human E3 ubiquitin
24	23	35.9	6	19	AAW56952	Enzyme inhibitor p
25	23	35.9	6	19	AAW56930	Enzyme inhibitor p
26	23	35.9	9	15	AAR57655	Peptide (155) inhi
27	23	35.9	9	15	AAR57661	Peptide (161) inhi
28	23	35.9	9	15	AAR59221	Peptide fragment (
29	23	35.9	9	21	AAB26335	Human CASB618 prot
30	23	35.9	9	21	AAB26356	Human CASB618 prot
31	23	35.9	9	22	AAB68094	Peptide derived fr
32	23	35.9	9	22	AAJ00492	Hepatitis C virus
33	23	35.9	9	22	AAJ00984	Hepatitis C virus
34	23	35.9	9	22	AAJ01799	Hepatitis C virus
35	23	35.9	9	22	AAJ02315	Hepatitis C virus
36	23	35.9	10	20	AAY47961	Immunogenic peptid
37	23	35.9	10	21	AAB26383	Human CASB618 prot
38	23	35.9	10	22	AAM42890	Mycoplasma genital
39	23	35.9	10	22	AAM42892	Mycoplasma genital
40	23	35.9	10	22	AAG96456	Human complementar
41	23	35.9	10	22	AAG97189	Human complementar
42	23	35.9	10	22	AAG97191	Human complementar
43	23	35.9	10	22	AAG85109	Saccharomyces cere
44	23	35.9	10	22	AAJ00493	Hepatitis C virus
45	23	35.9	10	22	AAJ01897	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAW47590
ID AAW47590 standard; peptide; 13 AA.
XX
AC AAW47590;
XX
DT 26-JUN-1998 (first entry)
XX
DE T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3.
XX
OS Homo sapiens.
XX
PN DE19625191-A1.
XX
PD 02-JAN-1998.
XX
PF 24-JUN-1996; 96DE-1025191.
XX
PR 24-JUN-1996; 96DE-1025191.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Schendel D;
XX
DR WPI; 1998-053442/06.
DR N-PSDB; AAV18707.
XX
PT Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX

```
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   13 AA;


  Query Match             100.0%;  Score 64;  DB 19;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CLVLSGSARQLTF 13
          |||||||||||||
Db      1 clvlsgsarqltf 13


RESULT   2
AAW47591
ID   AAW47591 standard; peptide; 12 AA.
XX
AC   AAW47591;
XX
DT   26-JUN-1998  (first entry)
XX
DE   T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3.
XX
OS   Homo sapiens.
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   N-PSDB; AAV18708.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   12 AA;


  Query Match             66.4%;  Score 42.5;  DB 19;  Length 12;
  Best Local Similarity   76.9%;  Pred. No. 0.1;
  Matches   10;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 CLVLSGSARQLTF 13
          ||  :|||||||
Db      1 cla-tgsarqltf 12


RESULT   3
AAG86220
ID   AAG86220 standard; Peptide; 10 AA.
XX
AC   AAG86220;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Saccharomyces cerevisiae peptide, SEQ ID NO: 1169.
XX
KW   Saccharomyces cerevisiae; complementary peptide; peptide identification;
KW   drug discovery; drug design.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200142276-A1.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04773.
XX
PR   13-DEC-1999;   99GB-0029471.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-367863/38.
XX
PT   Identifying complementary peptides by analysis of protein and
PT   nucleotide sequence databases, useful in drug design -
XX
PS   Example 3; Page 191; 488pp; English.
XX
CC   The invention relates to the identification of complementary peptides
CC   by analysis of protein and nucleotide sequence databases from higher
CC   eukaryotic genomes, excluding human and plants. The specific
CC   complementary peptides interact with their relevant target proteins
CC   encoded in the eukaryote genome. The peptides may be used as reagents
CC   and drugs for drug discovery and as lead ligands for drug design and
CC   development. The present sequence is a complementary peptide from
CC   Saccharomyces cerevisiae.
XX
SQ   Sequence   10 AA;


  Query Match             41.4%;  Score 26.5;  DB 22;  Length 10;
  Best Local Similarity   70.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 CLVL-SGSAR 9
          | || ||:||
Db      1 cfvlcsgtar 10


RESULT   4
AAG86310
ID   AAG86310 standard; Peptide; 10 AA.
XX
AC   AAG86310;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Saccharomyces cerevisiae peptide, SEQ ID NO: 1259.
XX
KW   Saccharomyces cerevisiae; complementary peptide; peptide identification;
KW   drug discovery; drug design.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200142276-A1.
XX
PD   14-JUN-2001.
XX
```

PF 13-DEC-2000; 2000WO-GB04773.
XX
PR 13-DEC-1999; 99GB-0029471.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-367863/38.
XX
PT Identifying complementary peptides by analysis of protein and
PT nucleotide sequence databases, useful in drug design -
XX
PS Example 3; Page 203; 488pp; English.
XX
CC The invention relates to the identification of complementary peptides
CC by analysis of protein and nucleotide sequence databases from higher
CC eukaryotic genomes, excluding human and plants. The specific
CC complementary peptides interact with their relevant target proteins
CC encoded in the eukaryote genome. The peptides may be used as reagents
CC and drugs for drug discovery and as lead ligands for drug design and
CC development. The present sequence is a complementary peptide from
CC Saccharomyces cerevisiae.
XX
SQ Sequence 10 AA;

Query Match 41.4%; Score 26.5; DB 22; Length 10;
Best Local Similarity 70.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 1 CLVL-SGSAR 9
Db 1 cfvlcsgtar 10

RESULT 5
AAB46917
ID AAB46917 standard; peptide; 9 AA.
XX
AC AAB46917;
XX
DT 04-MAY-2001 (first entry)
XX
DE Human SART-1 derived tumor antigenic peptide SEQ ID 3.
XX
KW Tumor antigenic peptide; interferon; IFN; antigen-specific T cell;
KW virucide; cytostatic; gene therapy; IFN-alpha; cytotoxic T cell;
KW viral infectious disease; SART-1; human.
XX
OS Homo sapiens.
XX
PN EP1074267-A1.
XX
PD 07-FEB-2001.
XX
PF 24-JUL-2000; 2000EP-0306263.
XX
PR 22-JUL-1999; 99JP-0207687.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Takasu H, Gotoh M, Yamaoka T;
XX
DR WPI; 2001-193144/20.
XX
PT Use of antigenic proteins, peptides, interferon or their encoding DNA,
PT in the manufacture of an agent for the induction of antigen-specific T
PT cells -
XX
PS Disclosure; Page 12; 25pp; English.
XX
CC This invention describes the novel use of interferons (IFNs) or DNAs
CC capable of expressing the interferons and/or antigenic proteins (AP),
CC antigenic peptides derived from the proteins or DNAs capable of
CC expressing the antigenic proteins or peptides, in the manufacture of
CC an agent for induction of antigen-specific T cells. The products of
CC the invention have virucide and cytostatic activity and can be used for
CC gene therapy or as inducers of antigen-specific T cells. The action of
CC interferon-alpha (IFN-alpha) in a system for inducing specific cytotoxic
CC T cell (CTL) by administering an antigenic peptide in an incomplete
CC Freund's adjuvant (IFA) emulsion preparation form was evaluated. IFNs (or
CC DNA encoding IFNs) are useful in the manufacture of a medicament for
CC inducing antigen-specific T cells in an individual who has been
CC administered with AP (or DNA encoding AP) or vice versa. The medicament
CC is useful for the treatment or prophylaxis of a tumor or a viral
CC infectious disease.
XX
SQ Sequence 9 AA;

Query Match 40.6%; Score 26; DB 22; Length 9;
Best Local Similarity 55.6%; Pred. No. 6.4e+05;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLSGSARQL 11
Db 1 vlsgsgksm 9

RESULT 6
AAG84200
ID AAG84200 standard; Peptide; 10 AA.
XX
AC AAG84200;
XX
DT 11-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana peptide ligand #840.
XX
KW Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
OS Arabidopsis thaliana.
XX
PN WO200142279-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04781.
XX
PR 13-DEC-1999; 99GB-0029469.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-381629/40.
XX
PT A set of peptide ligands for agricultural research and development of
PT therapeutic agents comprise specific complementary peptides to proteins
PT encoded by genes of plant genomes -
XX
PS Example 4; Page 148; 201pp; English.
XX
CC The present invention relates to a set of peptide ligands consisting of
CC specific complementary peptides to proteins encoded by genes of plant
CC genomes. The present sequence is one such peptide from Arabidopsis
CC thaliana. The peptides of the present invention are useful in an assay to
CC identify a peptide, especially a peptide pesticide or herbicide. The
CC peptides are also useful for tools for agricultural research and
CC development.
XX
SQ Sequence 10 AA;

Query Match 40.6%; Score 26; DB 22; Length 10;

```
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VLSGSA 8
           ||||||
Db       2 vlsgsa 7


RESULT   7
AAG84202
ID   AAG84202 standard; Peptide; 10 AA.
XX
AC   AAG84202;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Arabidopsis thaliana peptide ligand #842.
XX
KW   Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200142279-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04781.
XX
PR   13-DEC-1999;   99GB-0029469.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-381629/40.
XX
PT   A set of peptide ligands for agricultural research and development of
PT   therapeutic agents comprise specific complementary peptides to proteins
PT   encoded by genes of plant genomes -
XX
PS   Example 4; Page 148; 201pp; English.
XX
CC   The present invention relates to a set of peptide ligands consisting of
CC   specific complementary peptides to proteins encoded by genes of plant
CC   genomes. The present sequence is one such peptide from Arabidopsis
CC   thaliana. The peptides of the present invention are useful in an assay to
CC   identify a peptide, especially a peptide pesticide or herbicide. The
CC   peptides are also useful for tools for agricultural research and
CC   development.
XX
SQ   Sequence   10 AA;

  Query Match             40.6%;  Score 26;  DB 22;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VLSGSA 8
           ||||||
Db       3 vlsgsa 8


RESULT   8
AAW54287
ID   AAW54287 standard; Peptide; 11 AA.
XX
AC   AAW54287;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Cytohesin 1 PH domain randomised AB loop RM-18.
XX
KW   Pleckstrin homology domain; PH domain; peptide library;
KW   functional molecular surface; protein structural template;
KW   vaccine; gene therapy; cytohesin 1; human; plasmid pPHCY1;
KW   random mutagenesis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         4..8
FT                   /label= AB_loop
XX
PN   WO9745538-A1.
XX
PD   04-DEC-1997.
XX
PF   30-MAY-1997;   97WO-EP02840.
XX
PR   31-MAY-1996;   96EP-0108776.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bruhn H,  Funk M,  Henkel T,  Steipe B;
XX
DR   WPI; 1998-230215/20.
DR   N-PSDB; AAV26483.
XX
PT   Vectors used to produce PH domain-like peptide libraries - which are
PT   screened for therapeutically useful peptide(s), e.g. to produce
PT   vaccines
XX
PS   Example 2; Page 56; 137pp; English.
XX
CC   5 Peptide sequences (see AAW54286-90), respectively designated RM-1,
CC   RM-18, RM-21, RM-23 and RM-28, represent a peptide library derived
CC   from the AB loop (see AAW54291) of the Pleckstrin homology (PH)
CC   domain (see AAW54285) of human cytohesin 1.  They are encoded by
CC   5 in-frame mutant sequences (see AAV26482-86) generated by random
CC   mutagenesis of the DNA region (see AAV26487) encoding the AB loop.
CC   The average free energy of folding of mutant loop AB peptide RM-18
CC   is -27 kJ/mol, compared with -39 kJ/mol for the progenitor
CC   sequence.  The average free energy of folding of the 5 mutant
CC   peptides was -32.4 kJ/mol.  Randomisation does not compromise
CC   the structural integrity or the folding stability of the progenitor
CC   domain.  The invention provides vectors that are used for the
CC   production of PH domain-like peptide libraries, which can be
CC   screened to identify peptides that have desirable properties,
CC   especially novel binding or catalytic properties, and which may be
CC   of use in research or therapy, or as vaccines.  Novel synthetic
CC   protein structural templates for the generation, screening and
CC   evolution of functional molecular surfaces are provided.
XX
SQ   Sequence   11 AA;

  Query Match             40.6%;  Score 26;  DB 19;  Length 11;
  Best Local Similarity   62.5%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 GSARQLTF 13
           | ||:||:
Db       3 ghareltw 10


RESULT   9
AAY48993
ID   AAY48993 standard; Peptide; 9 AA.
XX
AC   AAY48993;
XX
DT   10-DEC-1999  (first entry)
XX
DE   Membrane dipeptidase-binding lymph node homing peptide #67.
XX
KW   Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;
```

KW prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;
KW membrane dipeptidase.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946284-A2.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1999; 99WO-US05284.
XX
PR 13-MAR-1998; 98US-0042107.
PR 26-FEB-1999; 99US-0042107.
XX
PA (BURN-) BURNHAM INST.
XX
PI Rajotte D, Pasqualini R, Ruoslahti EI;
XX
DR WPI; 1999-571717/48.
XX
PT New peptides which selectively home to organs or tissues, used for,
PT e.g. identifying target ligands and for therapy of pathological
PT conditions -
XX
PS Example 6; Page 154; 193pp; English.
XX
CC The present invention describes peptides that selectively home to a
CC tissue or organ. The peptides can be used for identifying an organ
CC or tissue, for identifying a target molecule expressed by an organ or
CC tissue or for treating an organ or tissue pathology, where the organ or
CC tissue is selected from prostate, lung, skin, retina, pancreas, gut,
CC ovary, adrenal gland, liver, and lymph node. The peptide bind to the
CC membrane dipeptidase (MDP). AAY48618 to AAY49066 represent sequences
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 25; DB 20; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVLS 5
|||||
Db 3 clvls 7

RESULT 10
AAB45862
ID AAB45862 standard; Protein; 9 AA.
XX
AC AAB45862;
XX
DT 21-MAR-2001 (first entry)
XX
DE Human tumor-associated antigen C42 protein immunogenic fragment #9.
XX
KW Tumor-associated antigen; C42; human; immunogenic; cancer therapy;
KW cytostatic; immunotherapy; vaccine; lung; breast; esophagus.
XX
OS Homo sapiens.
XX
PN DE19924199-A1.
XX
PD 30-NOV-2000.
XX
PF 27-MAY-1999; 99DE-1024199.
XX
PR 27-MAY-1999; 99DE-1024199.
XX
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
XX
PI Adolf G, Heider K, Koenig U, Sommergruber W, Adolf GR, Heider K;
PI Koenig U, Sommergruber W;
XX
DR WPI; 2001-062549/08.
XX
PT Tumor-associated antigen C42, and related nucleic acids and antibodies,
PT useful in immunotherapy of cancer and for diagnosis -
XX
PS Claim 5; Page 25; 40pp; German.
XX
CC This invention describes a novel tumor-associated antigen C42 which has
CC cytostatic activity. The invention also describes (a) immunogenic protein
CC fragments or peptides (I) derived from C42; (b) a pharmaceutical
CC composition for parenteral, topical, oral or local administration
CC containing at least 1 C42 or (I); (c) an isolated DNA (II) encoding C42
CC or its fragments; (d) a recombinant DNA (IIa) that contains (II); and
CC (e) antibodies (Ab) raised against C42 or (I). C42 and its fragments
CC induce a humoral immune response and, when presented by major
CC histocompatibility complex molecules, a cellular immune response. C42,
CC and its fragments, are used for immunotherapy of cancers that express C42
CC (particularly of lung, breast and esophagus), to raise specific
CC antibodies (Ab) and for diagnosis, e.g. to detect induction of an immune
CC response and for optimization of treatment. Ab are used therapeutically
CC (e.g. when conjugated to a cytotoxin) or for diagnosis or monitoring of
CC cancers that express C42. Nucleic acid (II) that encodes C42 can also be
CC used for immunotherapy and cells that express C42 as antitumor vaccines.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 25; DB 22; Length 9;
Best Local Similarity 55.6%; Pred. No. 6.4e+05;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLVLSGSAR 9
|||| |::
Db 1 clvldvssk 9

RESULT 11
AAG96480
ID AAG96480 standard; Peptide; 10 AA.
XX
AC AAG96480;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 2674.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 425; 646pp; English.

```
XX
CC   The invention relates to a set of complementary peptide ligands
CC   generated from the human genome. The complementary peptides
CC   interact with their relevant target proteins encoded in the human
CC   genome. They can be used as reagents in drug discovery and as lead
CC   ligands to facilitate drug design and development. The present
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;

  Query Match             39.1%;  Score 25;  DB 22;  Length 10;
  Best Local Similarity   83.3%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 SGSARQ 10
           :|||||
Db       4 agsarq 9


RESULT  12
AAG96482
ID   AAG96482 standard; Peptide; 10 AA.
XX
AC   AAG96482;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 2676.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04776.
XX
PR   13-DEC-1999;   99GB-0029464.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-408419/43.
XX
PT   A set of peptide ligands consisting of specific complementary peptides
PT   to proteins encoded by genes of the human genome, useful in an assay
PT   for screening and identifying of one or more novel peptides which are
PT   drug candidates or pro-drugs -
XX
PS   Example 4; Page 425; 646pp; English.
XX
CC   The invention relates to a set of complementary peptide ligands
CC   generated from the human genome. The complementary peptides
CC   interact with their relevant target proteins encoded in the human
CC   genome. They can be used as reagents in drug discovery and as lead
CC   ligands to facilitate drug design and development. The present
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;

  Query Match             39.1%;  Score 25;  DB 22;  Length 10;
  Best Local Similarity   83.3%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 SGSARQ 10
           :|||||
Db       5 agsarq 10


RESULT  13
AAW32678
ID   AAW32678 standard; peptide; 12 AA.
XX
AC   AAW32678;
XX
DT   30-JAN-1998  (first entry)
XX
DE   Human platelet glycoprotein Ib/IX complex epitope mimotope.
XX
KW   Human platelet glycoprotein Ib/IX complex epitope; mimotope; probe;
KW   monoclonal antibody C-34; anti mimotope; binding site; antibody;
KW   platelet adhesion; aggregation; agglutination; von Willebrand factor;
KW   anti thrombotic drug; ristocetin.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9718236-A1.
XX
PD   22-MAY-1997.
XX
PF   08-NOV-1996;   96WO-US17882.
XX
PR   13-NOV-1995;   95US-0556597.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Lyle VA,  Miller JL;
XX
DR   WPI; 1997-289227/26.
XX
PT   Peptide which mimics human platelet glycoprotein Ib/IX complex
PT   epitope - also anti-mimotope molecule capable of binding to the
PT   peptide and useful to modulate platelet adhesion, aggregation or
PT   agglutination
XX
PS   Claim 3; Page 97; 115pp; English.
XX
CC   The present sequence represents a mimotope which functionally mimics a
CC   binding site for a monoclonal antibody (MAb) which recognises an epitope
CC   within the human platelet glycoprotein (gp) Ib/IX complex. More
CC   specifically the MAb is C-34. The mimotope peptide can be used for
CC   raising antibodies, as probes to search for anti-mimotopes and to
CC   neutralise the inhibitory activity of those antibodies which recognise
CC   the epitope which is mimicked by the peptide. Anti-mimotopes can be used
CC   to modulate the adhesion, aggregation or agglutination of platelets by
CC   affecting von Willebrand factor interaction with the platelets through
CC   the gp Ib/IX receptor, e.g. as an anti-thrombotic drug which inhibits
CC   the ristocetin induced aggregation of platelets.
XX
SQ   Sequence   12 AA;

  Query Match             39.1%;  Score 25;  DB 18;  Length 12;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           |||: |
Db       4 clvmgg 9


RESULT  14
AAW71808
ID   AAW71808 standard; peptide; 12 AA.
XX
AC   AAW71808;
XX
DT   08-DEC-1998  (first entry)
XX
DE   Mimotope capable of binding to monoclonal antibody C-34 #57.
```

XX
KW Mimotope; monoclonal antibody C-34; human; glycoprotein Ib/IX complex;
KW binding site; platelet epitope; neutralise; aggregation; inhibition;
KW von Willebrand factor interaction; adhesion; agglutination.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US5817748-A.
XX
PD 06-OCT-1998.
XX
PF 17-MAR-1995; 95US-0406330.
XX
PR 17-MAR-1995; 95US-0406330.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Lyle VA, Miller JL;
XX
DR WPI; 1998-556458/47.
XX
PT Peptides that mimic platelet epitope - and neutralise
PT aggregation-inhibiting antibody
XX
PS Disclosure; Column 5; 26pp; English.
XX
CC The present invention provides peptides which functionally mimic a
CC binding site for a monoclonal antibody, where the monoclonal antibody
CC recognises an epitope within the human glycoprotein Ib/IX complex.
CC Peptides AAW71752 to AAW71832 are mimotopes which are capable of binding
CC to monoclonal antibody C-34. The invention also provides an isolated
CC molecule capable of binding the mimotopes, where the molecule can be
CC an antibody, a second peptide, a carbohydrate, a DNA molecule, an RNA
CC molecule, or other naturally or chemically synthesised molecules. These
CC molecules are called anti-mimotopes and can act as anti-thrombotic
CC drugs. A method has also been described for modulating the adhesion,
CC aggregation, or agglutination of platelets, where the method comprises
CC selecting platelets and exposing them to an anti-mimotope molecule. The
CC exposure affects von Willebrand factor interaction with platelets
CC through the glycoprotein Ib/IX receptor, which modulates the adhesion,
CC aggregation, or agglutination of the platelets. The peptide mimotopes
CC mimic platelet epitopes.
XX
SQ Sequence 12 AA;

Query Match 39.1%; Score 25; DB 19; Length 12;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLVLSG 6
||| :|
Db 4 clvmgg 9

RESULT 15
AAB84372
ID AAB84372 standard; peptide; 12 AA.
XX
AC AAB84372;
XX
DT 22-AUG-2001 (first entry)
XX
DE Splice-variance region of a human transferrin.
XX
KW Human; transferrin; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; aging; organ transplant.
XX
OS Homo sapiens.
XX
PN WO200146254-A1.
XX
PD 28-JUN-2001.
XX
PF 21-DEC-2000; 2000WO-US34769.
XX
PR 23-DEC-1999; 99US-0171595.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Shi Y, Choi GH;
XX
DR WPI; 2001-381910/40.
XX
PT Isolated nucleic acid molecule encoding a human tranferrin protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; Page 11; 290pp; English.
XX
CC The present sequence is a splice-variance region of a transferrin.
CC Transferrin polypeptides and polynucleotides are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. Disorders which are
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast
CC or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities.
XX
SQ Sequence 12 AA;

Query Match 39.1%; Score 25; DB 22; Length 12;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLVLSGSA 8
|: | | |
Db 3 cialkgea 10

Search completed: June 6, 2002, 06:23:03
Job time: 337 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:52 ; Search time 12.85 Seconds
(without alignments)
24.711 Million cell updates/sec

Title: US-08-881-509-6
Perfect score: 64
Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 88308

Minimum DB seq length: 0
Maximum DB seq length: 13

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	39.1	9	4	US-09-258-754-363	Sequence 363, App
2	25	39.1	9	4	US-09-042-107-363	Sequence 363, App
3	25	39.1	12	2	US-08-406-330-57	Sequence 57, Appl
4	25	39.1	12	2	US-08-556-597-57	Sequence 57, Appl
5	24	37.5	12	2	US-08-487-675-2	Sequence 2, Appli
6	24	37.5	12	3	US-08-904-263A-2	Sequence 2, Appli
7	23	35.9	13	2	US-08-479-846-4	Sequence 4, Appli
8	23	35.9	13	2	US-08-454-418A-4	Sequence 4, Appli
9	23	35.9	13	3	US-08-089-397A-5	Sequence 5, Appli
10	23	35.9	13	5	PCT-US93-04399-4	Sequence 4, Appli
11	22	34.4	6	1	US-08-201-046A-4	Sequence 4, Appli
12	22	34.4	8	1	US-08-526-710-32	Sequence 32, Appl
13	22	34.4	8	3	US-08-862-855-32	Sequence 32, Appl
14	22	34.4	8	4	US-09-226-985-32	Sequence 32, Appl
15	22	34.4	8	4	US-09-227-906-32	Sequence 32, Appl
16	22	34.4	9	2	US-08-482-651-10	Sequence 10, Appl
17	22	34.4	9	2	US-08-482-651-25	Sequence 25, Appl
18	22	34.4	9	4	US-08-660-092-13	Sequence 13, Appl
19	22	34.4	9	4	US-08-660-092-28	Sequence 28, Appl
20	22	34.4	9	6	5217869-105	Patent No. 5217869
21	22	34.4	12	1	US-07-834-902-6	Sequence 6, Appli
22	22	34.4	12	1	US-08-018-994-6	Sequence 6, Appli
23	22	34.4	12	1	US-08-294-675A-6	Sequence 6, Appli
24	22	34.4	12	3	US-09-188-579-3	Sequence 3, Appli
25	22	34.4	12	4	US-09-315-444-3	Sequence 3, Appli
26	21	32.8	8	2	US-08-669-284B-32	Sequence 32, Appl
27	21	32.8	9	2	US-08-482-651-8	Sequence 8, Appli
28	21	32.8	9	2	US-08-482-651-35	Sequence 35, Appl
29	21	32.8	9	4	US-08-660-092-11	Sequence 11, Appl
30	21	32.8	9	4	US-08-660-092-38	Sequence 38, Appl
31	21	32.8	9	4	US-08-660-092-59	Sequence 59, Appl
32	21	32.8	9	4	US-09-510-738A-42	Sequence 42, Appl
33	21	32.8	9	4	US-09-510-738A-58	Sequence 58, Appl
34	21	32.8	9	4	US-09-510-738A-99	Sequence 99, Appl
35	21	32.8	10	1	US-08-250-789A-4	Sequence 4, Appli
36	21	32.8	10	1	US-08-346-333-61	Sequence 61, Appl
37	21	32.8	10	2	US-08-968-542C-20	Sequence 20, Appl
38	21	32.8	10	3	US-08-925-002-7	Sequence 7, Appli
39	21	32.8	10	4	US-09-306-593-3	Sequence 3, Appli
40	21	32.8	10	4	US-09-230-548-3	Sequence 3, Appli
41	21	32.8	10	5	PCT-US91-07506-61	Sequence 61, Appl
42	21	32.8	11	2	US-08-748-021-22	Sequence 22, Appl
43	21	32.8	11	2	US-08-771-602D-36	Sequence 36, Appl
44	21	32.8	11	3	US-08-904-446A-9	Sequence 9, Appli
45	21	32.8	11	3	US-08-974-297-22	Sequence 22, Appl

ALIGNMENTS

```
RESULT   1
US-09-258-754-363
; Sequence 363, Application US/09258754
; Patent No. 6174687
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT: Pasqualini, Renata
;  APPLICANT: Rajotte, Daniel
;  TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using
;  TITLE OF INVENTION: Membrane Dipeptidase
;  FILE REFERENCE: P-LJ 3443
;  CURRENT APPLICATION NUMBER: US/09/258,754
;  CURRENT FILING DATE: 1999-02-26
;  EARLIER APPLICATION NUMBER: 09/042,107
;  EARLIER FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 452
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 363
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-258-754-363


  Query Match           39.1%;  Score 25;  DB 4;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVLS 5
             |||||
Db       3 CLVLS 7


RESULT   2
US-09-042-107-363
; Sequence 363, Application US/09042107
; Patent No. 6232287
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT: Pasqualini, Renata
;  TITLE OF INVENTION: Molecules that Home to Various Selected Organs or
;  TITLE OF INVENTION: Tissues
;  FILE REFERENCE: P-LJ 2892
;  CURRENT APPLICATION NUMBER: US/09/042,107
;  CURRENT FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 436
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 363
```

```
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-042-107-363


  Query Match             39.1%;  Score 25;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLVLS 5
           |||||
Db       3 CLVLS 7


RESULT   3
US-08-406-330-57
; Sequence 57, Application US/08406330
; Patent No. 5817748
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Jonathan L.
;    APPLICANT:  Lyle, Vicki A.
;    TITLE OF INVENTION:  MIMOTOPES AND ANTI-MIMOTOPES OF
;    TITLE OF INVENTION:  HUMAN PLATELET GLYCOPROTEIN Ib/IX
;    NUMBER OF SEQUENCES:  81
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,330
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  20884/100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1636
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-406-330-57


  Query Match             39.1%;  Score 25;  DB 2;  Length 12;
  Best Local Similarity   66.7%;  Pred. No. 72;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           |||: |
Db       4 CLVMGG 9


RESULT   4
US-08-556-597-57
; Sequence 57, Application US/08556597
; Patent No. 5877155
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Jonathan L.
;    APPLICANT:  Lyle, Vicki A.
;    TITLE OF INVENTION:  MIMOTOPES AND ANTI-MIMOTOPES OF
;    TITLE OF INVENTION:  HUMAN PLATELET GLYCOPROTEIN Ib/IX
;    NUMBER OF SEQUENCES:  173
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,597
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/406,330
;      FILING DATE:  17-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  20884/101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1636
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-556-597-57


  Query Match             39.1%;  Score 25;  DB 2;  Length 12;
  Best Local Similarity   66.7%;  Pred. No. 72;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           |||: |
Db       4 CLVMGG 9


RESULT   5
US-08-487-675-2
; Sequence 2, Application US/08487675
; Patent No. 5869064
;  GENERAL INFORMATION:
;    APPLICANT:  LINDAHL, Gunnar
;    APPLICANT:  STALHAMMAR-CARLEMALM, Margaretha
;    APPLICANT:  STENBERG, Lars
;    TITLE OF INVENTION:  PROTEIN RIB, A CELL SURFACE PROTEIN
;    TITLE OF INVENTION:  THAT CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP
;    TITLE OF INVENTION:  B STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE
;    TITLE OF INVENTION:  PROTEIN, REAGENT KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH, LLP
;      STREET:  8110 Gatehouse Road, Suite 500 East
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;       SOFTWARE:
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/487,675
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/SE94/00246
;       FILING DATE:  21-MAR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/SE93/00234
;       FILING DATE:  19-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Svensson, Leonard R.
;       REGISTRATION NUMBER:  30,330
;       REFERENCE/DOCKET NUMBER:  552-114P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  N-terminal
;     ORIGINAL SOURCE:
;       ORGANISM:  Streptococcus group B
;       STRAIN:  BM110
US-08-487-675-2


  Query Match             37.5%;  Score 24;  DB 2;  Length 12;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VLSGSA 8
           |:||||
Db       3 VISGSA 8


RESULT   6
US-08-904-263A-2
; Sequence 2, Application US/08904263A
; Patent No. 6015889
;  GENERAL INFORMATION:
;    APPLICANT:  LINDAHL, GUNNAR
;    APPLICANT:  STALHAMMAR-CARLEMALM,
;    APPLICANT:  MARGARETHA
;    APPLICANT:  STENBERG, LARS
;    TITLE OF INVENTION:  PROTEIN RIB, A CELL SURFACE PROTEIN THAT
;    TITLE OF INVENTION:  CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP B
;    TITLE OF INVENTION:  STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE PROTEIN,
;    TITLE OF INVENTION:  REAGENT KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH AND BIRCH, LLP
;      STREET:  PO BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,263A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SVENSSON, LEONARD R
;      REGISTRATION NUMBER:  30,330
;      REFERENCE/DOCKET NUMBER:  552-119P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-904-263A-2


  Query Match             37.5%;  Score 24;  DB 3;  Length 12;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VLSGSA 8
           |:||||
Db       3 VISGSA 8


RESULT   7
US-08-479-846-4
; Sequence 4, Application US/08479846
; Patent No. 5874534
;  GENERAL INFORMATION:
;    APPLICANT:  Vegeto, Elisabetta
;    APPLICANT:  McDonnell, Donald P.
;    APPLICANT:  O'Malley, Bert W.
;    APPLICANT:  Schrader, William T.
;    APPLICANT:  Tsai, Ming-Jer
;    TITLE OF INVENTION:  Mutated Steroid Hormone Receptors,
;    TITLE OF INVENTION:  Methods for Their Use and Molecular Switch for Gene
;    TITLE OF INVENTION:  Therapy
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,846
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/939,246
;      FILING DATE:  02-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5505
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5151
;      TELEFAX:  713/651-5246
;      TELEX:  762829
```

```
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  US 07/882,771
;      FILING DATE:  14-MAY-1992
US-08-479-846-4


  Query Match             35.9%;  Score 23;  DB 2;  Length 13;
  Best Local Similarity   42.9%;  Pred. No. 2e+02;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGS 7
           |::|| :
Db       2 CMILSNN 8


RESULT   8
US-08-454-418A-4
; Sequence 4, Application US/08454418A
; Patent No. 5935934
;  GENERAL INFORMATION:
;    APPLICANT:  Vegeto, Elisabetta
;    APPLICANT:  McDonnell, Donald P.
;    APPLICANT:  O'Malley, Bert W.
;    APPLICANT:  Schrader, William T.
;    APPLICANT:  Tsai, Ming-Jer
;    TITLE OF INVENTION:  MUTATED STEROID HORMONE RECEPTORS,
;    TITLE OF INVENTION:  METHODS FOR THEIR USE AND MOLECULAR
;    TITLE OF INVENTION:  SWITCH FOR GENE THERAPY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,418A
;      FILING DATE:  May 30, 1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/939,246
;      FILING DATE:  September 2, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkman, Charles S.
;      REGISTRATION NUMBER:  38,077
;      REFERENCE/DOCKET NUMBER:  212/295
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Peptide
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  US 07/882,771
;      FILING DATE:  May 14, 1992
US-08-454-418A-4


  Query Match             35.9%;  Score 23;  DB 2;  Length 13;
  Best Local Similarity   42.9%;  Pred. No. 2e+02;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGS 7
           |::|| :
Db       2 CMILSNN 8


RESULT   9
US-08-089-397A-5
; Sequence 5, Application US/08089397A
; Patent No. 6086880
;  GENERAL INFORMATION:
;    APPLICANT:  SABARA, MARTA I.J.
;    APPLICANT:  FRENCHICK, PATRICK J.
;    APPLICANT:  POTTER, ANDREW A.
;    APPLICANT:  IJAZ, MOHAMMAD K.
;    APPLICANT:  GILCHRIST, JAMES E.
;    APPLICANT:  REDMOND, MARK J.
;    TITLE OF INVENTION:  ROTAVIRUS VACCINES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/089,397A
;      FILING DATE:  07-JUL-1993
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ADLER, REID G.
;      REGISTRATION NUMBER:  30,988
;      REFERENCE/DOCKET NUMBER:  29311-20003.03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-089-397A-5


  Query Match             35.9%;  Score 23;  DB 3;  Length 13;
  Best Local Similarity   37.5%;  Pred. No. 2e+02;
  Matches    3;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 GSARQLTF 13
           |::||: :
Db       2 GASRQIVY 9
```

```
RESULT 10
PCT-US93-04399-4
; Sequence 4, Application PC/TUS9304399
;  GENERAL INFORMATION:
;     APPLICANT:  Vegeto, Elisabetta
;     APPLICANT:  McDonnell, Donald P.
;     APPLICANT:  O'Malley, Bert W.
;     APPLICANT:  Schrader, William T.
;     APPLICANT:  Tsai, Ming-Jer
;     TITLE OF INVENTION:  Mutated Steroid Hormone Receptors,
;     TITLE OF INVENTION:  Methods for Their Use and Molecular Switch for Gene
;     TITLE OF INVENTION:  Therapy
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fulbright & Jaworski
;       STREET:  1301 McKinney, Suite 5100
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  U.S.A.
;       ZIP:  77010-3095
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/04399
;       FILING DATE:  19930511
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/939,246
;       FILING DATE:  02-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Paul, Thomas D.
;       REGISTRATION NUMBER:  32,714
;       REFERENCE/DOCKET NUMBER:  D-5505
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713/651-5151
;       TELEFAX:  713/651-5246
;       TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  13 amino acids
;       TYPE:  AMINO ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  US 07/882,771
;       FILING DATE:  14-MAY-1992
PCT-US93-04399-4


  Query Match             35.9%;  Score 23;  DB 5;  Length 13;
  Best Local Similarity   42.9%;  Pred. No. 2e+02;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSGS 7
           |::|| :
Db       2 CMILSNN 8


RESULT 11
US-08-201-046A-4
; Sequence 4, Application US/08201046A
; Patent No. 5545719
;  GENERAL INFORMATION:
;     APPLICANT:  Shashoua, Victor E.
;     TITLE OF INVENTION:  NERVE GROWTH PEPTIDES AND USES THEREFOR
;     NUMBER OF SEQUENCES:  29
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;       STREET:  600 Atlantic Ave.
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/201,046A
;       FILING DATE:  24-FEB-1994
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Plumer, Elizabeth R.
;       REGISTRATION NUMBER:  36,637
;       REFERENCE/DOCKET NUMBER:  N0260/7013
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-720-3500
;       TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  internal
US-08-201-046A-4


  Query Match             34.4%;  Score 22;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           || :||
Db       1 CLPVSG 6


RESULT 12
US-08-526-710-32
; Sequence 32, Application US/08526710
; Patent No. 5622699
;  GENERAL INFORMATION:
;     APPLICANT:  Ruoslahti, Erkki
;     APPLICANT:  Pasqualini, Renata
;     TITLE OF INVENTION:  Method of Identifying Molecules That
;     TITLE OF INVENTION:  Home to a Selected Organ In Vivo
;     NUMBER OF SEQUENCES:  44
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Campbell and Flores
;       STREET:  4370 La Jolla Village Drive, Suite 700
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/526,710
;       FILING DATE:  11-SEP-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LJ 1779
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
```

```
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-526-710-32


  Query Match           34.4%;  Score 22;  DB 1;  Length 8;
  Best Local Similarity  66.7%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           | :|||
Db       1 CKLLSG 6


RESULT  13
US-08-862-855-32
; Sequence 32, Application US/08862855
; Patent No. 6068829
;  GENERAL INFORMATION:
;    APPLICANT:  Ruoslahti, Erkki
;    APPLICANT:  Pasqualini, Renata
;    TITLE OF INVENTION:  Method of Identifying Molecules That
;    TITLE OF INVENTION:  Home to a Selected Organ In Vivo
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/862,855
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,710
;      FILING DATE:  11-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/813,273
;      FILING DATE:  10-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-862-855-32


  Query Match           34.4%;  Score 22;  DB 3;  Length 8;
  Best Local Similarity  66.7%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           | :|||
Db       1 CKLLSG 6


RESULT  14
US-09-226-985-32
; Sequence 32, Application US/09226985
; Patent No. 6296832
;  GENERAL INFORMATION:
;    APPLICANT:  Ruoslahti, Erkki
;    APPLICANT:  Pasqualini, Renata
;    TITLE OF INVENTION:  Molecules That Home to a Selected Organ In Vivo
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/226,985
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,710
;      FILING DATE:  11-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/813,273
;      FILING DATE:  10-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/862,855
;      FILING DATE:  23-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 3423
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-226-985-32


  Query Match           34.4%;  Score 22;  DB 4;  Length 8;
  Best Local Similarity  66.7%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLVLSG 6
           | :|||
Db       1 CKLLSG 6


RESULT  15
US-09-227-906-32
; Sequence 32, Application US/09227906
; Patent No. 6306365
;  GENERAL INFORMATION:
;    APPLICANT:  Ruoslahti, Erkki
;    APPLICANT:  Pasqualini, Renata
;    TITLE OF INVENTION:  Method of Identifying Molecules That
;    TITLE OF INVENTION:  Home to a Selected Organ In Vivo
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
```

; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-32

Query Match 34.4%; Score 22; DB 4; Length 8;
Best Local Similarity 66.7%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 CLVLSG 6
           | :|||
Db       1 CKLLSG 6
```

Search completed: June 6, 2002, 06:23:22
Job time: 330 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:39:13 ; Search time 2149.71 Seconds
(without alignments)
350.445 Million cell updates/sec

Title: US-08-881-509-7
Perfect score: 36
Sequence: 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	100.0	36	6	A93133	A93133 Sequence 7
2	36	100.0	1318	9	HSTCRJUNC	X98410 H.sapiens m
3	29.6	82.2	1341	6	A93127	A93127 Sequence 1
4	28.6	79.4	716	9	S63879	S63879 TCR V alpha
5	28	77.8	39	9	HSU30448	U30448 Human isola
6	28	77.8	51	9	AF043886	AF043886 Homo sapi
7	27.6	76.7	39	6	A93131	A93131 Sequence 5
8	27.4	76.1	39	9	HSU30428	U30428 Human isola
9	27.4	76.1	51	9	AF043875	AF043875 Homo sapi
10	27.4	76.1	51	9	AF043877	AF043877 Homo sapi
11	27.4	76.1	51	9	AF043884	AF043884 Homo sapi
12	27.4	76.1	63	9	HSU30390	U30390 Human isola
13	27.4	76.1	411	9	HUMTCVJCC	M97706 Human T-cel
14	27.4	76.1	645	9	HSTCRAJ3	X02886 Human gene
15	27.4	76.1	71153	9	HUAE000662	AE000662 Homo sapi
16	27.4	76.1	97630	9	HUMTCRADCV	M94081 Human Tcr-C
17	27.4	76.1	175053	2	AC023226	AC023226 Homo sapi
18	27	75.0	48	9	HSU27254	U27254 Human isola
19	26.8	74.4	87	9	HUMTCRACG	L42801 Homo sapien
20	26.4	73.3	42	9	HSA235208	AJ235208 Homo sapi
21	26.4	73.3	51	9	AF043888	AF043888 Homo sapi
22	26.4	73.3	51	9	AF043890	AF043890 Homo sapi
23	26.4	73.3	183	9	HUMTCAZA	M17666 Human T-cel
24	26.4	73.3	222	9	AF333590	AF333590 Homo sapi
25	26	72.2	51	9	AF043874	AF043874 Homo sapi
26	26	72.2	51	9	AF043876	AF043876 Homo sapi
27	26	72.2	51	9	AF043892	AF043892 Homo sapi
28	26	72.2	51	9	AF043894	AF043894 Homo sapi
29	26	72.2	98	9	S69140	S69140 Homo sapien
30	26	72.2	432	9	HUMIGTCACA	M81774 Homo sapien
31	25.4	70.6	214	9	HSU40776	U40776 Human T-cel
32	25	69.4	234	9	AF041171	AF041171 Homo sapi
33	24.2	67.2	42	9	HSTCELL14	X81549 H.sapiens r
34	24.2	67.2	51	9	HSU14029	U14029 Human T cel
35	24.2	67.2	246	9	HSTCARAVB	Z22966 H.sapiens m
c 36	23.2	64.4	140531	9	AC046129	AC046129 Homo sapi
37	23.2	64.4	196629	2	AC016101	AC016101 Homo sapi
38	23.2	64.4	333458	2	AC087242	AC087242 Homo sapi
39	22.2	61.7	165253	2	OSJN00048	AL606615 Oryza sat
40	21.8	60.6	177686	2	AC023442	AC023442 Homo sapi
41	21.8	60.6	181925	2	AC020710	AC020710 Homo sapi
42	21.8	60.6	197864	9	AC021956	AC021956 Homo sapi
c 43	21.6	60.0	9157	10	AF295529	AF295529 Mus muscu
c 44	21.6	60.0	30933	2	AC109364	AC109364 Mus muscu
c 45	21.6	60.0	64695	2	AC101555	AC101555 Mus muscu

ALIGNMENTS

RESULT 1
A93133
LOCUS A93133 36 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent EP0816496.
ACCESSION A93133
VERSION A93133.1 GI:6741522
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 36)
AUTHORS Schendel,D.J.
TITLE T-cells specific for kidney carcinoma
JOURNAL Patent: EP 0816496-A 7 07-JAN-1998;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES Location/Qualifiers
source 1. .36
/organism="unidentified"
/db_xref="taxon:32644"
CDS <1. .>36
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69532.1"

```
                     /db_xref="GI:6741523"
                     /translation="CLATGSARQLTF"
BASE COUNT        6 a      11 c      8 g      11 t
ORIGIN


  Query Match             100.0%;  Score 36;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-05;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          |||||||||||||||||||||||||||||||||||
Db      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36


RESULT   2
HSTCRJUNC
LOCUS       HSTCRJUNC               1318 bp    mRNA    linear   PRI 08-JAN-1997
DEFINITION  H.sapiens mRNA for rearranged TCR junctional sequences.
ACCESSION   X98410
VERSION     X98410.1  GI:1770560
KEYWORDS    J gene; junction; T cell receptor; TCR junctional sequence; V gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1318)
  AUTHORS   Jantzer,P.U. and Schendel,D.J.
  TITLE     Tumor-infiltrating lymphocytes recognizing spontaneously arising
            renal cell carcinomas express T cell receptors characteristic of a
            secondary immune response
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1318)
  AUTHORS   Jantzer,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology,
            Goethestr. 31, D- 80336 Munich, FRG
FEATURES             Location/Qualifiers
     source          1. .1318
                     /organism="Homo sapiens"
                     /isolate="patients 22 and 26"
                     /db_xref="taxon:9606"
                     /rearranged
     CDS             <1. .>1318
                     /note="V gene/J gene junction"
                     /codon_start=1
                     /product="TCR junctional sequence"
                     /protein_id="CAA67057.1"
                     /db_xref="GI:1770561"
                     /translation="CLVGGSARQLTFGXCLATGSARQLTFGXCLVLSGSARQLTFGXC
                     LVLSGSARQLTFGXCLVLSGSARQLTFGXCLVGGSARQLTFGXCLVGGSARQLTFGXC
                     LVSGSARQLTFGXCLDSGSARQLTFGXCLDSGSARQLTFGXCLPSGSARQLTFGXCLA
                     TGSARQLTFGXCLVGGDTDKLIFGXCLVLSGSARQLTFGXCLVLSGSARQLTFGXCLV
                     VSGSARQLTFGXCLALGGSARQLTFGXCLALAGSARQLTFGXCLAPSGSARQLTFGXC
                     LVAGGGNTPLVFGXCLVGGSARQLTFGXCLVGGSARQLTFGXCLVGRSARQLTFGXCL
                     ATGSARQLTFGXCLVGAGGYQKVTFGXCLVLSGSARQLTFGXCLVAPGGATNKLIFGX
                     CLVVLGGSQGNLIFGXCLVGPNNAGNMLTFGXCLVATNQGRNCSDLW"
BASE COUNT      206 a    329 c    352 g    344 t     87 others
ORIGIN


  Query Match             100.0%;  Score 36;  DB 9;  Length 1318;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-05;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          |||||||||||||||||||||||||||||||||||
Db     43 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 78


RESULT   3
A93127
LOCUS       A93127                  1341 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0816496.
ACCESSION   A93127
VERSION     A93127.1  GI:6741516
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1341)
  AUTHORS   Schendel,D.J.
  TITLE     T-cells specific for kidney carcinoma
  JOURNAL   Patent: EP 0816496-A 1 07-JAN-1998;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .1341
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     sig_peptide     1. .54
     CDS             1. .804
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB69529.1"
                     /db_xref="GI:6741517"
                     /translation="MRQVARVIVFLTLSTLSLAKTTQPISMDSYEGQEVNITCSHNNI
                     ATNDYITWYQQFPSQGPRFIIQGYKTKVTNEVASLFIPADRKSSTLSLPRVSLSDTAV
                     YYCLVGGSARQLTFGSGTQLTVLPDIQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNV
                     SQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPE
                     SSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS"
     mat_peptide     55. .801
                     /product="unnamed"
BASE COUNT      331 a    365 c    294 g    351 t
ORIGIN


  Query Match             82.2%;  Score 29.6;  DB 6;  Length 1341;
  Best Local Similarity   88.9%;  Pred. No. 0.016;
  Matches   32;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          |||||||    ||||||||||||||||||||||||
Db    313 TGCCTCGTGGGTGGTTCTGCAAGGCAACTGACCTTT 348


RESULT   4
S63879
LOCUS       S63879                  716 bp    DNA     linear   PRI 04-MAY-2000
DEFINITION  TCR V alpha =T cell receptor variable alpha chain [human, MT-ALL,
            Genomic Mutant, 716 nt].
ACCESSION   S63879
VERSION     S63879.1  GI:238692
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Griesinger,F., Jansen,B. and Kersey,J.H.
  TITLE     Differentiation in mature T lymphoid leukemia cells is unstable and
            reversible to myeloid cells, without the involvement of a common
            stem cell
  JOURNAL   J. Immunol. 147 (10), 3336-3341 (1991)
  MEDLINE   92043697
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 63879] from the original journal article.
            This sequence comes from Figure 4.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .716
                     /gene="TCR V alpha"
                     /note="TCR V&agr"
     CDS             join(58. .109,340. .>716)
                     /gene="TCR V alpha"
```

```
                /note="This sequence comes from Figure 4; conceptual
                translation presented here differs from translation in
                publication"
                /codon_start=1
                /product="T cell receptor variable alpha chain"
                /protein_id="AAB20291.1"
                /db_xref="GI:238693"
                /translation="METLLEVLSGTLLWQLTWVRSQQPVQSPQAVILREGEDAVINCS
                SSKALYSVHWYRQKHGEAPVFLMILLKGGEQKGHEKISASFNEKKQQSSLYLTASQLS
                YSGTYFCGTASGSARQLTFGSGTQLTVLPDIQNPDPAVYQL"
BASE COUNT      197 a    159 c    177 g    183 t
ORIGIN


  Query Match             79.4%;  Score 28.6;  DB 9;  Length 716;
  Best Local Similarity   88.6%;  Pred. No. 0.046;
  Matches   31;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           ||   ||| ||||||||||||||||||||||||||
Db     616 GCACAGCTTCTGGTTCTGCAAGGCAACTGACCTTT 650


RESULT   5
HSU30448
LOCUS       HSU30448                 39 bp    mRNA    linear   PRI 01-AUG-1995
DEFINITION  Human isolate M94 T-cell receptor alpha V-J junction (TCR Valpha
            3/J alpha 22) mRNA, partial cds.
ACCESSION   U30448
VERSION     U30448.1  GI:915515
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
  TITLE     Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
  JOURNAL   Hum. Immunol. 37 (3), 178-184 (1993)
  MEDLINE   94064390
REFERENCE   2  (bases 1 to 39)
  AUTHORS   Hurwitz,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="Homo sapiens"
                     /isolate="M94"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
     gene            1. .39
                     /gene="TCR Valpha 3/J alpha 22"
     CDS             <1. .>39
                     /gene="TCR Valpha 3/J alpha 22"
                     /codon_start=1
                     /product="T-cell receptor alpha V-J junction"
                     /protein_id="AAA73621.1"
                     /db_xref="GI:915516"
                     /translation="FCVSSGSARQLTF"
BASE COUNT        5 a     11 c      8 g     15 t
ORIGIN


  Query Match             77.8%;  Score 28;  DB 9;  Length 39;
  Best Local Similarity   86.1%;  Pred. No. 0.095;
  Matches   31;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           ||  || |  |||||||||||||||||||||||||||
Db       4 TGTGTCTCCTCTGGTTCTGCAAGGCAACTGACCTTT 39


RESULT   6
AF043886
LOCUS       AF043886                 51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV314 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043886
VERSION     AF043886.1  GI:3859391
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV314"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1 [TCRAV16S1]
                     [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
                     and the 5' end of the TCRAJ/TCRBJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72681.1"
                     /db_xref="GI:3859392"
                     /translation="YFCAVRLTGSARQLTFG"
BASE COUNT       10 a     11 c     13 g     17 t
ORIGIN


  Query Match             77.8%;  Score 28;  DB 9;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 0.094;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 TACTGGTTCTGCAAGGCAACTGACCTTT 36
           ||||||||||||||||||||||||||||
Db      21 TACTGGTTCTGCAAGGCAACTGACCTTT 48


RESULT   7
A93131
LOCUS       A93131                   39 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 5 from Patent EP0816496.
ACCESSION   A93131
VERSION     A93131.1  GI:6741520
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Schendel,D.J.
  TITLE     T-cells specific for kidney carcinoma
  JOURNAL   Patent: EP 0816496-A 5 07-JAN-1998;
            BOEHRINGER MANNHEIM GMBH (DE)
```

```
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             <1. .>39
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB69531.1"
                     /db_xref="GI:6741521"
                     /translation="CLVLSGSARQLTF"
BASE COUNT        5 a     12 c      8 g     14 t
ORIGIN
```

```
  Query Match             76.7%;  Score 27.6;  DB 6;  Length 39;
  Best Local Similarity   88.2%;  Pred. No. 0.14;
  Matches   30;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           | ||  | ||||||||||||||||||||||||||||
Db       6 CGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
```

```
RESULT   8
HSU30428
LOCUS       HSU30428                  39 bp    mRNA    linear   PRI 01-AUG-1995
DEFINITION  Human isolate M74 T-cell receptor alpha V-J junction (TCR Valpha
            3/J alpha 22) mRNA, partial cds.
ACCESSION   U30428
VERSION     U30428.1  GI:915481
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
  TITLE     Restricted usage of T-cell receptor V alpha sequence and
            variable-joining pairs after normal T-cell development and bone
            marrow transplantation
  JOURNAL   Hum. Immunol. 37 (3), 178-184 (1993)
  MEDLINE   94064390
REFERENCE   2  (bases 1 to 39)
  AUTHORS   Hurwitz,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
            Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
            38101, USA
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="Homo sapiens"
                     /isolate="M74"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
     gene            1. .39
                     /gene="TCR Valpha 3/J alpha 22"
     CDS             <1. .>39
                     /gene="TCR Valpha 3/J alpha 22"
                     /codon_start=1
                     /product="T-cell receptor alpha V-J junction"
                     /protein_id="AAA73607.1"
                     /db_xref="GI:915482"
                     /translation="FCATSGSARQLTF"
BASE COUNT        6 a     10 c      8 g     15 t
ORIGIN
```

```
  Query Match             76.1%;  Score 27.4;  DB 9;  Length 39;
  Best Local Similarity   96.6%;  Pred. No. 0.18;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || ||||||||||||||||||||||||||||
Db      11 CTTCTGGTTCTGCAAGGCAACTGACCTTT 39
```

```
RESULT   9
AF043875
LOCUS       AF043875                  51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV2 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043875
VERSION     AF043875.1  GI:3859379
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
            synovial fluid of patients with rheumatoid arthritis
  JOURNAL   J. Immunol. 161 (8), 4428-4436 (1998)
  MEDLINE   98451502
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="AV2"
                     /tissue_type="synovial fluid"
                     /note="from rheumatoid arthritis patient CS-1, Identical
                     to AV325 [TCRAV1S2] [TCRAJ22]"
     gene            <1. .>51
                     /gene="TCRA"
     CDS             <1. .>51
                     /gene="TCRA"
                     /note="Contains the 3' end of the TCRAV, the n, and the 5'
                     end of the TCRAJ."
                     /codon_start=1
                     /product="T cell receptor alpha chain CDR3"
                     /protein_id="AAC72675.1"
                     /db_xref="GI:3859380"
                     /translation="YFCAVRSSGSARQLTFG"
BASE COUNT        7 a     12 c     14 g     18 t
ORIGIN
```

```
  Query Match             76.1%;  Score 27.4;  DB 9;  Length 51;
  Best Local Similarity   96.6%;  Pred. No. 0.18;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || ||||||||||||||||||||||||||||
Db      20 CTTCTGGTTCTGCAAGGCAACTGACCTTT 48
```

```
RESULT  10
AF043877
LOCUS       AF043877                  51 bp    mRNA    linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens patient CS-1 clone AV4 T cell receptor alpha chain
            CDR3 (TCRA) mRNA, partial cds.
ACCESSION   AF043877
VERSION     AF043877.1  GI:3859383
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
  TITLE     Selective accumulation of related CD4+ T cell clones in the
```

synovial fluid of patients with rheumatoid arthritis
JOURNAL J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE 98451502
REFERENCE 2 (bases 1 to 51)
AUTHORS Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and Research Center, 1400 Jackson St., Denver, CO 80206, USA

FEATURES Location/Qualifiers
source 1. .51
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV4"
/tissue_type="synovial fluid"
/note="from rheumatoid arthritis patient CS-1 [TCRAV1S3] [TCRAJ22]"
gene <1. .>51
/gene="TCRA"
CDS <1. .>51
/gene="TCRA"
/note="Contains the 3' end of the TCRAV, the n, and the 5' end of the TCRAJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72677.1"
/db_xref="GI:3859384"
/translation="YFCAVRPSGSARQLTFG"

BASE COUNT 9 a 12 c 13 g 17 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 51;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| ||||||||||||||||||||||||||||
Db 20 CTTCTGGTTCTGCAAGGCAACTGACCTTT 48

RESULT 11
AF043884
LOCUS AF043884 51 bp mRNA linear PRI 11-NOV-1998
DEFINITION Homo sapiens patient CS-1 clone AV325 T cell receptor alpha chain CDR3 (TCRA) mRNA, partial cds.
ACCESSION AF043884
VERSION AF043884.1 GI:3859387
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 51)
AUTHORS Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Selective accumulation of related CD4+ T cell clones in the synovial fluid of patients with rheumatoid arthritis
JOURNAL J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE 98451502
REFERENCE 2 (bases 1 to 51)
AUTHORS Striebich,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and Research Center, 1400 Jackson St., Denver, CO 80206, USA

FEATURES Location/Qualifiers
source 1. .51
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV325"
/tissue_type="synovial fluid"
/note="from rheumatoid arthritis patient CS-1 [TCRAV1S2] [TCRAJ22]"
gene <1. .>51
/gene="TCRA"
CDS <1. .>51
/gene="TCRA"
/note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n, and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72679.1"
/db_xref="GI:3859388"
/translation="YFCAVRSSGSARQLTFG"

BASE COUNT 7 a 12 c 14 g 18 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 51;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| ||||||||||||||||||||||||||||
Db 20 CTTCTGGTTCTGCAAGGCAACTGACCTTT 48

RESULT 12
HSU30390
LOCUS HSU30390 63 bp mRNA linear PRI 01-AUG-1995
DEFINITION Human isolate M101 T-cell receptor alpha V-J junction (TCR Valpha 5/J alpha 22) mRNA, partial cds.
ACCESSION U30390
VERSION U30390.1 GI:915415
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 63)
AUTHORS Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE Restricted usage of T-cell receptor V alpha sequence and variable-joining pairs after normal T-cell development and bone marrow transplantation
JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE 94064390
REFERENCE 2 (bases 1 to 63)
AUTHORS Hurwitz,J.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN 38101, USA

FEATURES Location/Qualifiers
source 1. .63
/organism="Homo sapiens"
/isolate="M101"
/db_xref="taxon:9606"
/tissue_type="blood"
gene 1. .63
/gene="TCR Valpha 5/J alpha 22"
CDS <1. .>63
/gene="TCR Valpha 5/J alpha 22"
/codon_start=1
/product="T-cell receptor alpha V-J junction"
/protein_id="AAA73579.1"
/db_xref="GI:915416"
/translation="TYLCALDPRTVSSGSARQLTF"

BASE COUNT 11 a 20 c 12 g 20 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 63;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| ||||||||||||||||||||||||||||
Db 35 CTTCTGGTTCTGCAAGGCAACTGACCTTT 63

```
RESULT  13
HUMTCVJCC
LOCUS       HUMTCVJCC                411 bp    mRNA    linear   PRI 14-JAN-1995
DEFINITION  Human T-cell receptor (V alpha 22.1, J alpha C, C alpha 1) mRNA.
ACCESSION   M97706
VERSION     M97706.1  GI:292803
KEYWORDS    T-cell receptor; T-cell receptor alpha.
SOURCE      Homo sapiens (individual_isolate AL61.270) cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hurley,C.K., Steiner,N., Wagner,A., Geiger,M.J., Eckels,D.D. and
            Rosen-Bronson,S.
  TITLE     Nonrandom T cell receptor usage in the allorecognition of HLA-DR1
            microvariation
  JOURNAL   J. Immunol. 150 (4), 1314-1324 (1993)
  MEDLINE   93163556
COMMENT     *haplotype = DRB1*0101
            Reference [J. Immunol. (1993) In press] reports bases 324 to 393.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Homo sapiens"
                     /isolate="AL61.270"
                     /db_xref="taxon:9606"
                     /map="14q11.2"
                     /haplotype="DRB1*0101"
                     /cell_type="alloproliferative T lymphocyte clone"
     sig_peptide     <1. .53
                     /gene="TCRA"
                     /note="G00-120-404; does not fit consensus"
                     /product="T-cell receptor alpha"
     gene            1. .411
                     /gene="TCRA"
     V_region        54. .323
                     /gene="TCRA"
                     /standard_name="V alpha 22.1"
                     /note="G00-120-404; putative"
                     /citation=[1]
                     /product="T-cell receptor alpha"
     J_segment       324. .393
                     /gene="TCRA"
                     /standard_name="J alpha C"
                     /note="Includes CDR3 region and J segment; G00-120-404;
                     putative"
                     /citation=[1]
                     /product="T-cell receptor alpha"
     C_region        394. .411
                     /gene="TCRA"
                     /standard_name="Constant alpha 1"
                     /note="G00-120-404; putative"
                     /citation=[1]
                     /product="T-cell receptor alpha"
BASE COUNT      107 a    106 c     92 g    106 t
ORIGIN


  Query Match           76.1%;  Score 27.4;  DB 9;  Length 411;
  Best Local Similarity   96.6%;  Pred. No. 0.16;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || |||||||||||||||||||||||||||
Db     334 CTTCTGGTTCTGCAAGGCAACTGACCTTT 362


RESULT  14
HSTCRAJ3
LOCUS       HSTCRAJ3                 645 bp    DNA     linear   PRI 11-APR-1995
DEFINITION  Human gene for T-cell receptor alpha chain J region.
ACCESSION   X02886
VERSION     X02886.1  GI:36928
KEYWORDS    joining region; T-cell receptor; T-cell receptor alpha.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Yoshikai,Y., Clark,S.P., Taylor,S., Sohn,U., Wilson,B.I.,
            Minden,M.D. and Mak,T.W.
  TITLE     Organization and sequences of the variable, joining and constant
            region genes of the human T-cell receptor alpha-chain
  JOURNAL   Nature 316 (6031), 837-840 (1985)
  MEDLINE   85296333
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    377. .385
                     /note="nonamer recombination signal"
     misc_feature    398. .404
                     /note="heptamer recombination signal"
     J_segment       405. .467
                     /note="J segment (405 is 2nd base in codon) (467 is 1st
                     base in codon)"
BASE COUNT      201 a    107 c    149 g    188 t
ORIGIN


  Query Match           76.1%;  Score 27.4;  DB 9;  Length 645;
  Best Local Similarity   96.6%;  Pred. No. 0.16;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || |||||||||||||||||||||||||||
Db     408 CTTCTGGTTCTGCAAGGCAACTGACCTTT 436


RESULT  15
HUAE000662
LOCUS       HUAE000662               71153 bp    DNA     linear   PRI 15-JUN-2001
DEFINITION  Homo sapiens T-cell receptor alpha delta locus from bases 1000498
            to 1071650 (section 5 of 5) of the Complete Nucleotide Sequence.
ACCESSION   AE000662 AE000521
VERSION     AE000662.1  GI:2358068
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71153)
  AUTHORS   Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
            Howard,S., Shan,W., Deshpande,P. and Hood,L.
  TITLE     The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
            organization, sequence, and evolution of 97.6 kb of DNA
  JOURNAL   Genomics 19 (3), 478-493 (1994)
  MEDLINE   94245236
  REMARK    This citation covers from bases 966383-1064019
REFERENCE   2  (bases 1 to 71153)
  AUTHORS   Boysen,C., Simon,M.I. and Hood,L.
  TITLE     Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
            bacterial artificial chromosome clones
  JOURNAL   Genome Res. 7 (4), 330-338 (1997)
  MEDLINE   97264339
   PUBMED   9110172
REFERENCE   3  (bases 1 to 71153)
  AUTHORS   Boysen,C., Inyoul,L., Smith,T.M., Smit,A., Wang,K., Rowen,L. and
            Hood,L.
  TITLE     T-Cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
  JOURNAL   Unpublished
  REMARK    This citation covers bases 1-983545 and bases 1064020-1071650
REFERENCE   4  (bases 1 to 71153)
  AUTHORS   Boysen,C., Lee,I., Smith,T.M., Smit,A., Kai,W., Lee,R. and Leroy,H.
  TITLE     Direct Submission
```

```
JOURNAL    Submitted (20-JUL-1997) Department of Molecular Biotechnology,
           University of Washington, Box 357730, Seattle, Washington 98195,
           USA
REMARK     Complete nucleotide sequence of the human T-cell receptor alpha
           delta locus
COMMENT    Sequencing method: high redundancy shotgun.  Interspersed Repeats
           were identified with RepeatMasker (available from
           http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
           sequence repeats were identified with sputnik (available from
           http://serac.mbt.washington.edu/~chrisa/software/sputnik.html).
FEATURES             Location/Qualifiers
     source          1. .71153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   2362. .2660
                     /rpt_family="AluSc"
     repeat_region   4274. .4574
                     /rpt_family="AluSx"
     repeat_region   9569. .9796
                     /rpt_family="MIR"
     repeat_region   10535. .10834
                     /rpt_family="AluY"
     repeat_region   10835. .11138
                     /rpt_family="AluJb"
     repeat_region   complement(13900. .14190)
                     /rpt_family="AluSq"
     repeat_region   complement(20506. .20810)
                     /rpt_family="AluY"
     repeat_region   20928. .21223
                     /rpt_family="AluSx"
     repeat_region   23083. .23335
                     /rpt_family="MIR"
     repeat_region   25320. .25618
                     /rpt_family="AluSq"
     repeat_region   complement(27167. .27240)
                     /rpt_family="MIR"
     repeat_region   27306. .27428
                     /rpt_family="MIR2"
     repeat_region   complement(30158. .30403)
                     /rpt_family="MIR"
     repeat_region   33727. .33878
                     /rpt_family="MIR"
     repeat_region   complement(33920. .34213)
                     /rpt_family="AluYa5"
     repeat_region   complement(35275. .35553)
                     /rpt_family="MIR2"
     repeat_region   complement(36848. .36886)
                     /rpt_family="L1MA3"
     repeat_region   39321. .39398
                     /rpt_family="MIR"
     repeat_region   40206. .40261
                     /rpt_family="MIR"
     repeat_region   complement(40433. .40659)
                     /rpt_family="MIR"
     repeat_region   complement(40683. .40848)
                     /rpt_family="AluSx"
     repeat_region   complement(41322. .41444)
                     /rpt_family="MIR"
     repeat_region   complement(45732. .45964)
                     /rpt_family="MIR"
     repeat_region   46927. .46992
                     /rpt_family="MIR2"
     repeat_region   complement(47281. .47388)
                     /rpt_family="MIR2"
     repeat_region   49066. .49073
                     /rpt_type=tandem
                     /rpt_unit=TAT
     repeat_region   49734. .50032
                     /rpt_family="AluY"
     repeat_region   51737. .51940
                     /rpt_family="MER3"
     repeat_region   complement(52892. .53047)
                     /rpt_family="MIR"
     repeat_region   54558. .54732
                     /rpt_family="MIR2"
     repeat_region   complement(54749. .55052)
                     /rpt_family="AluSq"
     repeat_region   55059. .55302
                     /rpt_family="MIR2"
     repeat_region   57276. .57570
                     /rpt_family="AluJo"
     repeat_region   59210. .59460
                     /rpt_family="MER46"
     repeat_region   complement(61706. .61769)
                     /rpt_family="MIR2"
     repeat_region   63506. .63580
                     /rpt_family="MIR"
     repeat_region   63695. .63862
                     /rpt_family="L1ME3"
     repeat_region   63931. .63951
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   64131. .64281
                     /rpt_family="MIR"
     repeat_region   65845. .66137
                     /rpt_family="AluSx"
     repeat_region   complement(66237. .66480)
                     /rpt_family="MIR2"
     repeat_region   complement(68187. .68457)
                     /rpt_family="AluJo"
     repeat_region   complement(68480. .68777)
                     /rpt_family="AluSq"
     repeat_region   complement(68778. .68837)
                     /rpt_family="FAM"
     repeat_region   69326. .69628
                     /rpt_family="AluSp"
     repeat_region   complement(69688. .69766)
                     /rpt_family="MER58A"
     repeat_region   69845. .69897
                     /rpt_family="MER5A"
     repeat_region   complement(69907. .70048)
                     /rpt_family="MER5A"
     repeat_region   complement(70060. .70225)
                     /rpt_family="MER5B"
BASE COUNT    20125 a  15457 c  16410 g  19161 t
ORIGIN


  Query Match             76.1%;  Score 27.4;  DB 9;  Length 71153;
  Best Local Similarity   96.6%;  Pred. No. 0.14;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || ||||||||||||||||||||||||||||
Db   28166 CTTCTGGTTCTGCAAGGCAACTGACCTTT 28194


Search completed: June  6, 2002, 07:39:15
Job time: 5378 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:54:29 ; Search time 2649.09 Seconds
(without alignments)
183.418 Million cell updates/sec

Title: US-08-881-509-7
Perfect score: 36
Sequence: 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	26.4	73.3	564	12	AQ221133	AQ221133 HS_3223_B
c	2	21.8	60.6	1060	12	CNS04W88	AL310049 Tetraodon
	3	21.6	60.0	284	9	AI715777	AI715777 UI-R-Y0-a
c	4	21.2	58.9	282	9	BB491342	BB491342 BB491342
	5	21.2	58.9	316	12	AZ492378	AZ492378 1M0326B02
c	6	21.2	58.9	1002	12	CNS079XE	AL435864 T7 end of
	7	21.2	58.9	1088	12	CNS0771C	AL432118 T7 end of
c	8	20.8	57.8	393	9	AA019771	AA019771 ze62e03.r
c	9	20.8	57.8	488	12	AQ423489	AQ423489 CITBI-E1-
c	10	20.8	57.8	506	12	AQ470385	AQ470385 CITBI-E1-
	11	20.8	57.8	667	10	BF304050	BF304050 601886829
c	12	20.8	57.8	852	10	BG479738	BG479738 602526929
c	13	20.6	57.2	440	12	AZ923465	AZ923465 4908.gf21
	14	20.6	57.2	527	10	BI534791	BI534791 397951 MA
c	15	20.6	57.2	667	9	BB639586	BB639586 BB639586
c	16	20.6	57.2	874	10	BF242338	BF242338 601880546
	17	20.4	56.7	120	10	BI423478	BI423478 949048D12
c	18	20.4	56.7	371	10	BF899667	BF899667 IL5-MT020
c	19	20.4	56.7	547	12	AQ049405	AQ049405 cLM-14c12
	20	20.4	56.7	619	12	BH062049	BH062049 RPCI-24-3
	21	20.4	56.7	1098	10	BI912745	BI912745 603176136
c	22	20.2	56.1	462	12	BH236190	BH236190 ATZKB19TF
	23	20.2	56.1	475	12	BH235936	BH235936 ATZKE49TR
c	24	20.2	56.1	586	12	BH236089	BH236089 ATZKE92TR
	25	20.2	56.1	618	9	BB615995	BB615995 BB615995
c	26	20.2	56.1	657	12	BH235920	BH235920 ATZKB30TF
	27	20.2	56.1	664	12	BH235999	BH235999 ATZKE32TF
c	28	20.2	56.1	694	12	BH235850	BH235850 ATZKF36TR
	29	20.2	56.1	939	10	BG366394	BG366394 HVSME1000
	30	20.2	56.1	960	9	AL581298	AL581298 AL581298
	31	20.2	56.1	985	12	CNS04DT3	AL286176 Tetraodon
c	32	20.2	56.1	1015	10	BE744953	BE744953 601576221
	33	20	55.6	219	10	BM191333	BM191333 dae17g04.
	34	20	55.6	283	9	AV171596	AV171596 AV171596
	35	20	55.6	288	9	AV147538	AV147538 AV147538
	36	20	55.6	301	9	AV136764	AV136764 AV136764
	37	20	55.6	330	10	BI314592	BI314592 dae14f12.
	38	20	55.6	348	9	AV110678	AV110678 AV110678
	39	20	55.6	357	10	T97080	T97080 ye50g08.r1
	40	20	55.6	359	10	BG513863	BG513863 dae03f08.
c	41	20	55.6	402	12	BH074362	BH074362 RPCI-24-2
	42	20	55.6	415	10	BG264799	BG264799 daa32g10.
	43	20	55.6	466	10	BG022611	BG022611 dab06f12.
	44	20	55.6	477	12	AQ432798	AQ432798 HS_5132_B
	45	20	55.6	501	10	BG022723	BG022723 dab07h06.

ALIGNMENTS

RESULT 1
AQ221133
LOCUS AQ221133 564 bp DNA linear GSS 19-SEP-1998
DEFINITION HS_3223_B2_E10_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3223 Col=20 Row=J, DNA sequence.
ACCESSION AQ221133
VERSION AQ221133.1 GI:3634746
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 564)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3223 row: J column: 20
Class: BAC ends
High quality sequence stop: 564.
FEATURES Location/Qualifiers
source 1. .564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3223 Col=20 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

```
BASE COUNT      160 a    102 c    124 g    171 t      7 others
ORIGIN


  Query Match             73.3%;  Score 26.4;  DB 12;  Length 564;
  Best Local Similarity   96.4%;  Pred. No. 2.4;
  Matches   27;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 TACTGGTTCTGCAAGGCAACTGACCTTT 36
           | ||||||||||||||||||||||||||
Db     423 TTCTGGTTCTGCAAGGCAACTGACCTTT 450


RESULT   2
CNS04W88/c
LOCUS       CNS04W88                1060 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence SP6 end of clone
            006M18 of library B from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL310049
VERSION     AL310049.1  GI:9542917
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1060)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1060)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="006M18"
                     /clone_lib="B"
                     /note="Genoscope sequence ID : C0AB006BG09B1-end : SP6"
BASE COUNT      264 a    250 c    231 g    292 t     23 others
ORIGIN


  Query Match             60.6%;  Score 21.8;  DB 12;  Length 1060;
  Best Local Similarity   78.8%;  Pred. No. 2e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 CTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           |||||| | ||||||| ||| || | | |||||
Db     625 CTCGCTCCGGGTTCTGAAAGCCATCAGTCCTTT 593


RESULT   3
```

```
AI715777
LOCUS       AI715777                284 bp    mRNA    linear   EST 10-JUN-1999
DEFINITION  UI-R-Y0-abx-a-05-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-abx-a-05-0-UI 3', mRNA sequence.
ACCESSION   AI715777
VERSION     AI715777.1  GI:5033030
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized Eye library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-Y0-abx-a-05-0-UI"
                     /clone_lib="UI-R-Y0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                     library is a subtracted library derived from an
                     individually-tagged normalized whole-eye (minus the lens)
                     library. The driver for the subtraction consisted of a
                     pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                     UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                     3-5 nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-Y0) was constructed as follows: PCR
                     amplified cDNA inserts from previous library clones from
                     which 3' ESTs had  been derived were used as a driver in a
                     hybridization with the normalized  whole-eye library in
                     the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-Y0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)
                     TAG_LIB=UI-R-Y0
                     TAG_TISSUE=Eye
                     TAG_SEQ=CATTG"
BASE COUNT       60 a     60 c     67 g     97 t
ORIGIN
```

```
  Query Match           60.0%;  Score 21.6;  DB 9;  Length 284;
  Best Local Similarity 75.0%;  Pred. No. 1.6e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           |||||| || || || |||    ||||||| |||||
Db     173 TGCCTCTCTGCTTGTGCTGTGCTGCAACTGTCCTTT 208


RESULT   4
BB491342/c
LOCUS       BB491342                282 bp    mRNA    linear   EST 25-JUL-2000
DEFINITION  BB491342 RIKEN full-length enriched, 13 days embryo stomach Mus
            musculus cDNA clone D530020L13 3', mRNA sequence.
ACCESSION   BB491342
VERSION     BB491342.1  GI:9449969
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D530020L13"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     stomach"
                     /tissue_type="stomach"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       79 a     52 c     50 g    100 t      1 others
ORIGIN


  Query Match           58.9%;  Score 21.2;  DB 9;  Length 282;
  Best Local Similarity 88.5%;  Pred. No. 2.3e+02;
  Matches   23;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       9 TACTGGTTCTGCAAGGCAACTGACCT 34
           |||| |||||||||||| ||||||| |
Db      50 TACTCGTTCTGCAAGGTAACTGACTT 25


RESULT   5
AZ492378
LOCUS       AZ492378                316 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0326B02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0326B02 R, DNA sequence.
ACCESSION   AZ492378
VERSION     AZ492378.1  GI:10665040
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0326  row: B  column: 02
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0326B02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
```

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 72 a 75 c 78 g 91 t
ORIGIN

Query Match 58.9%; Score 21.2; DB 12; Length 316;
Best Local Similarity 76.5%; Pred. No. 2.4e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy       1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           || || | ||||| |||| | ||||| |||| ||
Db     161 TGGCTTGTTACTGCTTCTTCCAGGCAGCTGATCT 194
```

RESULT 6
CNS079XE/c
LOCUS CNS079XE 1002 bp DNA linear GSS 08-JUL-2001
DEFINITION T7 end of clone XBB0AA001H06 of library XBB0AA from strain CBS 4732 of Pichia angusta, genomic survey sequence.
ACCESSION AL435864
VERSION AL435864.1 GI:12219277
KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 1002)
AUTHORS Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
REFERENCE 2 (bases 1 to 1002)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
REFERENCE 3 (bases 1 to 1002)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .1002
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="XBB0AA001H06"
/clone_lib="XBB0AA"
/note="end : T7"
misc_feature complement(<946. .>999)
/note="similar to P06834 [Dihydroxy-acetone synthase (DHAS)] [Pichia angusta]"
/evidence=not_experimental

BASE COUNT 209 a 226 c 281 g 281 t 5 others
ORIGIN

Query Match 58.9%; Score 21.2; DB 12; Length 1002;
Best Local Similarity 76.5%; Pred. No. 3.4e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| || ||||| ||||| | || ||| ||
Db     154 GCATCGCGACAGGTTCCGCAAGCCCACAGACTTT 121
```

RESULT 7
CNS0771C
LOCUS CNS0771C 1088 bp DNA linear GSS 08-JUL-2001
DEFINITION T7 end of clone BB0AA008E09 of library BB0AA from strain CBS 4732 of Pichia angusta, genomic survey sequence.
ACCESSION AL432118
VERSION AL432118.1 GI:12215532
KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 1088)
AUTHORS Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
REFERENCE 2 (bases 1 to 1088)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
REFERENCE 3 (bases 1 to 1088)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .1088
/organism="Pichia angusta"

```
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA008E09"
                     /clone_lib="BB0AA"
                     /note="end : T7"
BASE COUNT       293 a    335 c    236 g    223 t      1 others
ORIGIN


  Query Match             58.9%;  Score 21.2;  DB 12;  Length 1088;
  Best Local Similarity   76.5%;  Pred. No. 3.5e+02;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| || ||||| ||||| | || ||| ||
Db     259 GCATCGCGACAGGTTCCGCAAGCCCACAGACTTT 292


RESULT   8
AA019771/c
LOCUS       AA019771                393 bp    mRNA    linear   EST 30-JAN-1997
DEFINITION  ze62e03.r1 Soares retina N2b4HR Homo sapiens cDNA clone
            IMAGE:363580 5', mRNA sequence.
ACCESSION   AA019771
VERSION     AA019771.1  GI:1483107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2024    Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Homo sapiens"
                     /db_xref="GDB:1280283"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:363580"
                     /clone_lib="Soares retina N2b4HR"
                     /sex="male"
                     /tissue_type="retina"
                     /dev_stage="55 year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). The retinas were obtained from a 55 year old
                     Caucasian  and total cellular poly(A)+ RNA was extracted 6
                     hrs after their removal. The retina RNA  was kindly
                     provided by Roderick R. McInnes M.D. Ph.D. from the
                     University  of Toronto. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. "
BASE COUNT       132 a     82 c     74 g    103 t      2 others
ORIGIN


  Query Match             57.8%;  Score 20.8;  DB 9;  Length 393;
  Best Local Similarity   75.8%;  Pred. No. 3.6e+02;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACC 33
           | | ||   ||  |||||||||||| |||||||
Db     367 TTCTTCNGGACCAGTTCTGCAAGGCTACTGACC 335


RESULT   9
AQ423489/c
LOCUS       AQ423489                488 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  CITBI-E1-2568P10.TF CITBI-E1 Homo sapiens genomic clone 2568P10,
            DNA sequence.
ACCESSION   AQ423489
VERSION     AQ423489.1  GI:4481213
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2568P10.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2568P10"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT       153 a     91 c    126 g    118 t
ORIGIN


  Query Match             57.8%;  Score 20.8;  DB 12;  Length 488;
  Best Local Similarity   78.1%;  Pred. No. 3.9e+02;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 CTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
           ||| |||||| |||  | |||| ||||| |||
Db     251 CTCCCTACTGTTTCACCTAGGCCACTGAACTT 220


RESULT  10
AQ470385/c
LOCUS       AQ470385                506 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  CITBI-E1-2584F3.TF CITBI-E1 Homo sapiens genomic clone 2584F3, DNA
```

```
            sequence.
ACCESSION   AQ470385
VERSION     AQ470385.1  GI:4654039
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2584F3.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2584F3"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      160 a     91 c    127 g    128 t
ORIGIN
```

```
  Query Match             57.8%;  Score 20.8;  DB 12;  Length 506;
  Best Local Similarity   78.1%;  Pred. No. 3.9e+02;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 CTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
           ||| |||||| |||  | |||| ||||| |||
Db     251 CTCCCTACTGTTTCACCTAGGCCACTGAACTT 220
```

```
RESULT  11
BF304050
LOCUS       BF304050                667 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  601886829F2 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4120807 5',
            mRNA sequence.
ACCESSION   BF304050
VERSION     BF304050.1  GI:11250727
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 667)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1001  row: j  column: 08
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1. .667
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4120807"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      173 a    169 c    174 g    151 t
ORIGIN
```

```
  Query Match             57.8%;  Score 20.8;  DB 10;  Length 667;
  Best Local Similarity   78.1%;  Pred. No. 4.3e+02;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 TCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           || | ||||||| || |||||||||  ||| |
Db     591 TCCCAACTGGTTATGTAAGGCAACTTCCCTCT 622
```

```
RESULT  12
BG479738/c
LOCUS       BG479738                852 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602526929F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4650702 5',
            mRNA sequence.
ACCESSION   BG479738
VERSION     BG479738.1  GI:13412017
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1432  row: i  column: 07
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4650702"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
```

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 219 a 185 c 177 g 271 t
ORIGIN

Query Match 57.8%; Score 20.8; DB 10; Length 852;
Best Local Similarity 78.1%; Pred. No. 4.6e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACC 33
|||| |||||||| || ||| | |||||||
Db 477 GCCTAGCTACTGGACCTCAAAGACCACTGACC 446

RESULT 13
AZ923465/c
LOCUS AZ923465 440 bp DNA linear GSS 01-APR-2001
DEFINITION 4908.gf21c06.s1 Saccharomyces cariocanus UFRJ 50791 Saccharomyces cariocanus genomic clone 4908.gf21c06.s1, DNA sequence.
ACCESSION AZ923465
VERSION AZ923465.1 GI:13494363
KEYWORDS GSS.
SOURCE Saccharomyces cariocanus.
ORGANISM Saccharomyces cariocanus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 440)
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
FEATURES Location/Qualifiers
source 1. .440
/organism="Saccharomyces cariocanus"
/strain="UFRJ 50791"
/db_xref="taxon:114526"
/clone="4908.gf21c06.s1"
/clone_lib="Saccharomyces cariocanus UFRJ 50791"
/note="Random genomic sequence "
BASE COUNT 132 a 69 c 86 g 153 t
ORIGIN

Query Match 57.2%; Score 20.6; DB 12; Length 440;
Best Local Similarity 74.3%; Pred. No. 4.5e+02;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
||| ||||| || |||| || | |||||||| |
Db 136 TGCAACGCTATTGACTCTGTAAAGAAACTGACCGT 102

RESULT 14
BI534791
LOCUS BI534791 527 bp mRNA linear EST 30-AUG-2001
DEFINITION 397951 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BI534791
VERSION BI534791.1 GI:15375899
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 527)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 123 row: G column: 14
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .527
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."
BASE COUNT 132 a 126 c 123 g 146 t
ORIGIN

Query Match 57.2%; Score 20.6; DB 10; Length 527;
Best Local Similarity 74.3%; Pred. No. 4.8e+02;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
|||||| || ||| |||||||| || ||| ||
Db 73 TGCCTCTATAATGGCTCTGCAAGACAGCTGTTTTT 107

RESULT 15
BB639586/c
LOCUS BB639586 667 bp mRNA linear EST 26-OCT-2001
DEFINITION BB639586 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730061C10 5', mRNA sequence.
ACCESSION BB639586
VERSION BB639586.1 GI:16474972
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 667)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .667
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730061C10"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 201 a 180 c 140 g 145 t 1 others
ORIGIN

Query Match 57.2%; Score 20.6; DB 9; Length 667;
Best Local Similarity 85.2%; Pred. No. 5.2e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      5 TCGCTACTGGTTCTGCAAGGCAACTGA 31
          ||  || ||||||||||||||| ||||
Db    124 TCATTAGTGGTTCTGCAAGGCACCTGA 98
```

Search completed: June 6, 2002, 06:54:33

Job time: 2696 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:00:23 ; Search time 333.99 Seconds
(without alignments)
185.062 Million cell updates/sec

Title: US-08-881-509-7
Perfect score: 36
Sequence: 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36	100.0	36	19	AAV18708	cDNA for T-cell re
	2	29.6	82.2	1341	19	AAV18705	cDNA for T-cell re
	3	27.6	76.7	39	19	AAV18707	cDNA for T-cell re
c	4	19.6	54.4	5482	23	ABL23904	Drosophila melanog
	5	19.4	53.9	133	22	ABA47504	Human breast cell
	6	19.4	53.9	133	22	ABA65394	Human foetal liver
	7	19.4	53.9	133	22	ABA32492	Probe #10958 for g
	8	19.4	53.9	133	22	AAK13803	Human brain expres
	9	19.4	53.9	133	22	AAK39548	Human bone marrow
	10	19.4	53.9	133	22	AAI20356	Probe #10289 for g
	11	19.4	53.9	133	22	AAI45561	Probe #14247 used
	12	19.4	53.9	133	22	AAI06058	Probe #6049 used t
c	13	19.4	53.9	264	22	ABA17256	Human nervous syst
c	14	19.4	53.9	265	22	ABA17254	Human nervous syst
c	15	19.4	53.9	265	22	ABA17255	Human nervous syst
	16	19.4	53.9	456	22	ABA42363	Human breast cell
	17	19.4	53.9	456	22	ABA52789	Human foetal liver
	18	19.4	53.9	456	22	ABA22574	Probe #1040 for ge
	19	19.4	53.9	456	22	AAK01043	Human brain expres
	20	19.4	53.9	456	22	AAK26502	Human bone marrow
	21	19.4	53.9	456	22	AAI11130	Probe #1063 for ge
	22	19.4	53.9	456	22	AAI32395	Probe #1081 used t
	23	19.4	53.9	456	22	AAI01048	Probe #1039 used t
	24	19	52.8	1001	21	AAH51555	Human FMO related
c	25	19	52.8	1473	22	AAF60759	Psudomonas sp ABC
	26	19	52.8	1632	24	AAS62328	cDNA sequence #115
	27	19	52.8	26016	19	AAV57272	Human flavin-conta
c	28	18.8	52.2	1456	22	AAI61190	Human polynucleoti
c	29	18.6	51.7	302	20	AAX29171	LHP polypeptide en
c	30	18.6	51.7	303	22	AAA89038	Mycobacterium tube
c	31	18.6	51.7	327	18	AAT91509	Mycobacterium tube
c	32	18.6	51.7	327	18	AAT91445	Mycobacterium tube
c	33	18.6	51.7	327	19	AAV64492	M. tuberculosis im
c	34	18.6	51.7	327	19	AAV44384	Mycobacterium tube
c	35	18.6	51.7	327	20	AAZ19294	M. tuberculosis an
c	36	18.6	51.7	327	20	AAZ19082	M. tuberculosis re
c	37	18.6	51.7	396	18	AAT91526	Mycobacterium tube
c	38	18.6	51.7	396	18	AAT91460	Mycobacterium tube
c	39	18.6	51.7	396	19	AAV64509	M. tuberculosis im
c	40	18.6	51.7	396	19	AAV44400	Mycobacterium tube
c	41	18.6	51.7	396	20	AAZ19310	M. tuberculosis an
c	42	18.6	51.7	396	20	AAZ19098	M. tuberculosis re
c	43	18.6	51.7	460	22	ABA42575	Human breast cell
c	44	18.6	51.7	460	22	ABA53004	Human foetal liver
c	45	18.6	51.7	460	22	ABA22779	Probe #1245 for ge

ALIGNMENTS

```
RESULT   1
AAV18708
ID   AAV18708 standard; cDNA; 36 BP.
XX
AC   AAV18708;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1..36
FT                   /*tag=  a
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;    96DE-1025191.
XX
PR   24-JUN-1996;    96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
```

```
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47591.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence encodes the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence 36 BP; 6 A; 11 C; 8 G; 11 T; 0 other;

  Query Match             100.0%;  Score 36;  DB 19;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 3e-06;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          |||||||||||||||||||||||||||||||||||
Db      1 tgcctcgctactggttctgcaaggcaactgaccttt 36


RESULT   2
AAV18705
ID   AAV18705 standard; cDNA; 1341 BP.
XX
AC   AAV18705;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor alpha-chain.
XX
KW   Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..804
FT                   /*tag=  a
FT   sig_peptide     1..54
FT                   /*tag=  b
FT   mat_peptide     55..801
FT                   /*tag=  c
FT                   /product= T-cell_receptor_alpha-chain
FT
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47588.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Pages 11-13; 30pp; German.
XX
CC   The present sequence encodes the alpha-chain of a human T-cell
CC   receptor (TCR), which can be used in the diagnosis, monitoring,
CC   prevention and therapy of a tumour disease, specifically renal
CC   cell carcinoma.
XX
SQ   Sequence 1341 BP; 331 A; 365 C; 294 G; 351 T; 0 other;

  Query Match             82.2%;  Score 29.6;  DB 19;  Length 1341;
  Best Local Similarity   88.9%;  Pred. No. 0.0038;
  Matches   32;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          |||||||    |||||||||||||||||||||||||
Db    313 tgcctcgtgggtggttctgcaaggcaactgaccttt 348


RESULT   3
AAV18707
ID   AAV18707 standard; cDNA; 39 BP.
XX
AC   AAV18707;
XX
DT   26-JUN-1998  (first entry)
XX
DE   cDNA for T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1..39
FT                   /*tag=  a
FT
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   P-PSDB; AAW47590.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence encodes the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence 39 BP; 5 A; 12 C; 8 G; 14 T; 0 other;

  Query Match             76.7%;  Score 27.6;  DB 19;  Length 39;
  Best Local Similarity   88.2%;  Pred. No. 0.013;
  Matches   30;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 CCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
          | ||  | |||||||||||||||||||||||||||
Db      6 cgtcctttctggttctgcaaggcaactgaccttt 39


RESULT   4
ABL23904/c
ID   ABL23904 standard; DNA; 5482 BP.
```

XX
AC ABL23904;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23185.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 23185; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5482 BP; 1352 A; 1226 C; 1338 G; 1566 T; 0 other;

Query Match 54.4%; Score 19.6; DB 23; Length 5482;
Best Local Similarity 73.5%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCTT 35
||| | | |||||| ||| | ||||| || | ||
Db 37 GCCGCTCGACTGGTCCTGAATGGCAAGTGGCTTT 4

RESULT 5
ABA47504
ID ABA47504 standard; DNA; 133 BP.
XX
AC ABA47504;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #6199.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 4; SEQ ID NO 6199; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;
Best Local Similarity 79.3%; Pred. No. 63;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
|| |||| ||||||| | ||||| ||||
Db 46 gccactgcttctgcagtggaactgtcctt 74

RESULT 6
ABA65394
ID ABA65394 standard; DNA; 133 BP.
XX
AC ABA65394;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #13699.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.

```
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 13699; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

  Query Match             53.9%;  Score 19.4;  DB 22;  Length 133;
  Best Local Similarity   79.3%;  Pred. No. 63;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| ||||||| | ||||| ||||
Db      46 gccactgcttctgcagtggaactgtcctt 74


RESULT   7
ABA32492
ID   ABA32492 standard; DNA; 133 BP.
XX
AC   ABA32492;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #10958 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 4; SEQ ID No 10958; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

  Query Match             53.9%;  Score 19.4;  DB 22;  Length 133;
  Best Local Similarity   79.3%;  Pred. No. 63;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| ||||||| | ||||| ||||
Db      46 gccactgcttctgcagtggaactgtcctt 74


RESULT   8
AAK13803
ID   AAK13803 standard; DNA; 133 BP.
XX
AC   AAK13803;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 13794.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
```

PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 13794; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;
Best Local Similarity 79.3%; Pred. No. 63;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
|| |||| ||||||| | ||||| ||||
Db 46 gccactgcttctgcagtggaactgtcctt 74

RESULT 9
AAK39548
ID AAK39548 standard; DNA; 133 BP.
XX
AC AAK39548;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 14105.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 14105; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;
Best Local Similarity 79.3%; Pred. No. 63;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
|| |||| ||||||| | ||||| ||||
Db 46 gccactgcttctgcagtggaactgtcctt 74

RESULT 10
AAI20356
ID AAI20356 standard; DNA; 133 BP.
XX
AC AAI20356;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #10289 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 10289; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;
Best Local Similarity 79.3%; Pred. No. 63;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
|| |||| ||||||| | ||||| ||||
Db 46 gccactgcttctgcagtggaactgtcctt 74

```
RESULT 11
AAI45561
ID   AAI45561 standard; DNA; 133 BP.
XX
AC   AAI45561;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #14247 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 14247; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

  Query Match           53.9%;  Score 19.4;  DB 22;  Length 133;
  Best Local Similarity 79.3%;  Pred. No. 63;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| ||||||| | ||||| ||||
Db      46 gccactgcttctgcagtggaactgtcctt 74


RESULT 12
AAI06058
ID   AAI06058 standard; DNA; 133 BP.
XX
AC   AAI06058;
XX
DT   09-OCT-2001 (first entry)
XX
DE   Probe #6049 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 6049; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

  Query Match           53.9%;  Score 19.4;  DB 22;  Length 133;
  Best Local Similarity 79.3%;  Pred. No. 63;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| ||||||| | ||||| ||||
Db      46 gccactgcttctgcagtggaactgtcctt 74


RESULT 13
ABA17256/c
ID   ABA17256 standard; DNA; 264 BP.
XX
AC   ABA17256;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9587.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
```

XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9587; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 264 BP; 51 A; 65 C; 81 G; 67 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 264;
Best Local Similarity 79.3%; Pred. No. 73;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
            || |||| ||||||| | ||||| ||||
Db      54 GCCACTGCTTCTGCAGTGGAACTGCCCTT 26
```

RESULT 14
ABA17254/c
ID ABA17254 standard; DNA; 265 BP.
XX
AC ABA17254;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9585.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9585; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 265 BP; 51 A; 67 C; 81 G; 66 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 265;
Best Local Similarity 79.3%; Pred. No. 73;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
|| |||| ||||||| | ||||| ||||
Db 54 GCCACTGCTTCTGCAGTGGAACTGCCCTT 26

RESULT 15
ABA17255/c
ID ABA17255 standard; DNA; 265 BP.
XX
AC ABA17255;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9586.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9586; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 265 BP; 51 A; 67 C; 81 G; 66 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 265;
Best Local Similarity 79.3%; Pred. No. 73;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35
           || |||| ||||||| | ||||| ||||
Db      54 GCCACTGCTTCTGCAGTGGAACTGCCCTT 26
```

Search completed: June 6, 2002, 07:00:24
Job time: 3047 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:38:15 ; Search time 78.65 Seconds
(without alignments)
112.432 Million cell updates/sec

Title: US-08-881-509-7
Perfect score: 36
Sequence: 1 TGCCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19.4	53.9	15297	4	US-09-817-180-3	Sequence 3, Appli
	2	18.8	52.2	518	4	US-09-060-756-628	Sequence 628, App
c	3	18.6	51.7	327	4	US-08-818-112-46	Sequence 46, Appl
c	4	18.6	51.7	327	4	US-08-818-111-46	Sequence 46, Appl
c	5	18.6	51.7	327	4	US-09-056-556-46	Sequence 46, Appl
c	6	18.6	51.7	396	4	US-08-818-112-116	Sequence 116, App
c	7	18.6	51.7	396	4	US-08-818-111-111	Sequence 111, App
c	8	18.6	51.7	396	4	US-09-056-556-116	Sequence 116, App
	9	18.6	51.7	657	1	US-08-076-011-3	Sequence 3, Appli
	10	18.6	51.7	710	4	US-08-998-416-850	Sequence 850, App
c	11	18.6	51.7	1616	4	US-08-818-112-112	Sequence 112, App
c	12	18.6	51.7	1616	4	US-08-818-111-107	Sequence 107, App
c	13	18.6	51.7	1616	4	US-09-056-556-112	Sequence 112, App
c	14	18.6	51.7	7676	4	US-09-056-556-213	Sequence 213, App
c	15	18.6	51.7	16885	1	US-08-390-878-16	Sequence 16, Appl
	16	18.6	51.7	17041	1	US-08-076-011-1	Sequence 1, Appli
c	17	18.6	51.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	18	18.6	51.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	19	18.2	50.6	2220	6	5171850-1	Patent No. 5171850
c	20	18	50.0	2340	4	US-09-450-852-3	Sequence 3, Appli
	21	18	50.0	29604	3	US-08-781-891-207	Sequence 207, App
c	22	17.8	49.4	1125	3	US-08-878-801-3	Sequence 3, Appli
c	23	17.8	49.4	2388	4	US-09-276-531-89	Sequence 89, Appl
	24	17.6	48.9	213	4	US-08-905-223-46	Sequence 46, Appl
	25	17.6	48.9	1101	3	US-08-746-883-3	Sequence 3, Appli
	26	17.6	48.9	1447	4	US-09-180-437-3	Sequence 3, Appli
	27	17.6	48.9	2243	3	US-09-084-079-4	Sequence 4, Appli
	28	17.6	48.9	2404	4	US-09-180-437-4	Sequence 4, Appli
c	29	17.4	48.3	6306	1	US-08-466-390-3	Sequence 3, Appli
c	30	17.4	48.3	6306	1	US-08-470-950-3	Sequence 3, Appli
c	31	17.4	48.3	6306	1	US-08-467-781-3	Sequence 3, Appli
c	32	17.4	48.3	6306	1	US-08-195-487-3	Sequence 3, Appli
c	33	17.4	48.3	6306	2	US-08-483-924-3	Sequence 3, Appli
c	34	17.4	48.3	6306	5	PCT-US93-06160-3	Sequence 3, Appli
c	35	17.4	48.3	13011	2	US-08-791-849A-14	Sequence 14, Appl
	36	17.4	48.3	49136	4	US-09-422-869-1	Sequence 1, Appli
c	37	17.2	47.8	99	4	US-08-905-223-95	Sequence 95, Appl
	38	17.2	47.8	209	4	US-09-085-199B-31	Sequence 31, Appl
	39	17.2	47.8	1164	4	US-09-085-199B-1	Sequence 1, Appli
c	40	17.2	47.8	1207	2	US-08-667-809B-1	Sequence 1, Appli
c	41	17.2	47.8	1424	2	US-08-915-972A-21	Sequence 21, Appl
c	42	17.2	47.8	1424	2	US-09-177-909-21	Sequence 21, Appl
c	43	17.2	47.8	1424	4	US-09-056-105-21	Sequence 21, Appl
c	44	17.2	47.8	3266	1	US-08-369-043-3	Sequence 3, Appli
c	45	17.2	47.8	4406	1	US-08-369-043-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-817-180-3/c
; Sequence 3, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001183
;  CURRENT APPLICATION NUMBER: US/09/817,180
;  CURRENT FILING DATE:  2001-03-27
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 15297
;   TYPE: DNA
;   ORGANISM: Human
US-09-817-180-3


  Query Match             53.9%;  Score 19.4;  DB 4;  Length 15297;
  Best Local Similarity   79.3%;  Pred. No. 22;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5 TCGCTACTGGTTCTGCAAGGCAACTGACC 33
           ||| ||| |||||||||| ||||   ||||
Db    9087 TCGGTACAGGTTCTGCAGGGCAGGGGACC 9059


RESULT 2
US-09-060-756-628
; Sequence 628, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 628
;   LENGTH: 518
;   TYPE: DNA
```

; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-628

Query Match 52.2%; Score 18.8; DB 4; Length 518;
Best Local Similarity 74.2%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGAC 32
|||| ||||| || || ||||||| |||
Db 467 gcctggctacgggctcaacaaggcanaagac 497

RESULT 3
US-08-818-112-46/c
; Sequence 46, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-46

Query Match 51.7%; Score 18.6; DB 4; Length 327;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
||| ||| ||||| |||||| | | ||| | |
Db 131 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 99

RESULT 4
US-08-818-111-46/c
; Sequence 46, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-46

Query Match 51.7%; Score 18.6; DB 4; Length 327;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
||| ||| ||||| |||||| | | ||| | |
Db 131 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 99

RESULT 5
US-09-056-556-46/c
; Sequence 46, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TRE
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/056,556
;       FILING DATE:  07-APR-1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.457
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  327 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-056-556-46


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 327;
  Best Local Similarity   72.7%;  Pred. No. 19;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db     131 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 99


RESULT   6
US-08-818-112-116/c
; Sequence 116, Application US/08818112
; Patent No. 6290969
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonio
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND DIAGNOSIS OF TUBERCULOSIS
;    NUMBER OF SEQUENCES:  153
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,112
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.411C6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  116:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  396 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-818-112-116


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 396;
  Best Local Similarity   72.7%;  Pred. No. 20;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db      76 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 44


RESULT   7
US-08-818-111-111/c
; Sequence 111, Application US/08818111
; Patent No. 6338852
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonia
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,111
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.417C6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  111:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  396 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-818-111-111


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 396;
  Best Local Similarity   72.7%;  Pred. No. 20;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db      76 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 44


RESULT   8
US-09-056-556-116/c
```

```
; Sequence 116, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATM
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  116:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  396 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-116


  Query Match           51.7%;  Score 18.6;  DB 4;  Length 396;
  Best Local Similarity  72.7%;  Pred. No. 20;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db      76 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 44


RESULT  9
US-08-076-011-3
; Sequence 3, Application US/08076011
; Patent No. 5521069
;  GENERAL INFORMATION:
;    APPLICANT:  ONDA, Haruo
;    APPLICANT:  KIMURA, Chiharu
;    APPLICANT:  OHKUBO, Shoichi
;    TITLE OF INVENTION:  NOVEL DNA AND USE THEREOF
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN,
;      ADDRESSEE:  ROBERTS & CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,011
;      FILING DATE:  11-JUN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/047,246
;      FILING DATE:  13-APR-1993
;      NAME:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/741,676
;      FILING DATE:  07-AUG-1991
;      NAME:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNICK, DAVID S
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  41155-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  657 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-076-011-3


  Query Match           51.7%;  Score 18.6;  DB 1;  Length 657;
  Best Local Similarity  72.7%;  Pred. No. 23;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 CTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           ||| || ||||||||||  | | ||  || |||
Db     485 CTCCCTCCTGGTTCTGCGCGTCTACAAACTTTT 517


RESULT  10
US-08-998-416-850
; Sequence 850, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Meigs, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCXET NUMBER:  PF/5-30306/A/CGC1976
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8587
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  850:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  710 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  PAG1536UP
US-08-998-416-850


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 710;
  Best Local Similarity   72.7%;  Pred. No. 23;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 CTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
           |||| ||||||     |||||| |||| || ||
Db       7 CTCGGTACTGGAGAAACAAGGCTACTGTCCCTT 39


RESULT  11
US-08-818-112-112/c
; Sequence 112, Application US/08818112
; Patent No. 6290969
;  GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Skeiky, Yasir A.W.
;     APPLICANT:  Dillon, Davin C.
;     APPLICANT:  Campos-Neto, Antonio
;     APPLICANT:  Houghton, Raymond
;     APPLICANT:  Vedvick, Thomas S.
;     APPLICANT:  Twardzik, Daniel R.
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;     TITLE OF INVENTION:  AND DIAGNOSIS OF TUBERCULOSIS
;     NUMBER OF SEQUENCES:  153
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/818,112
;       FILING DATE:  13-MAR-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.411C6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  112:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1616 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-818-112-112


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 1616;
  Best Local Similarity   72.7%;  Pred. No. 29;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db    1353 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 1321


RESULT  12
US-08-818-111-107/c
; Sequence 107, Application US/08818111
; Patent No. 6338852
;  GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Skeiky, Yasir A.W.
;     APPLICANT:  Dillon, Davin C.
;     APPLICANT:  Campos-Neto, Antonia
;     APPLICANT:  Houghton, Raymond
;     APPLICANT:  Vedvick, Thomas S.
;     APPLICANT:  Twardzik, Daniel R.
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;     NUMBER OF SEQUENCES:  148
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/818,111
;       FILING DATE:  13-MAR-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.417C6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  107:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1616 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-818-111-107


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 1616;
  Best Local Similarity   72.7%;  Pred. No. 29;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db    1353 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 1321


RESULT  13
US-09-056-556-112/c
; Sequence 112, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G.
;     APPLICANT:  Skeiky, Yasir A.W.
;     APPLICANT:  Dillon, Davin C.
```

```
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATM
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  112:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1616 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-112


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 1616;
  Best Local Similarity   72.7%;  Pred. No. 29;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db    1353 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 1321


RESULT  14
US-09-056-556-213/c
; Sequence 213, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATM
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  213:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7676 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-213


  Query Match             51.7%;  Score 18.6;  DB 4;  Length 7676;
  Best Local Similarity   72.7%;  Pred. No. 42;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db    6475 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 6443


RESULT  15
US-08-390-878-16/c
; Sequence 16, Application US/08390878
; Patent No. 5700683
;  GENERAL INFORMATION:
;    APPLICANT:  Stover, Charles K.
;    APPLICANT:  Mahairas, Gregory G.
;    TITLE OF INVENTION:  VIRULENCE-ATTENUATING GENETIC DELETIONS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower, 20th
;      STREET:  Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,878
;      FILING DATE:  17-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  15371A-17
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/543/9600
;      TELEFAX:  415/543/5043
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16885 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-390-878-16


  Query Match             51.7%;  Score 18.6;  DB 1;  Length 16885;
  Best Local Similarity   72.7%;  Pred. No. 50;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
           ||| ||| |||||  |||||| | | ||| | |
Db    4473 GCCGCGCCACTGGCCCTGCAACGAACCTGCCGT 4441
```

Search completed: June 6, 2002, 07:38:20
Job time: 5322 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:16:22 ; Search time 27.21 Seconds
(without alignments)
42.377 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	67.2	63	2	A41912	T-cell receptor al
2	40	65.6	384	2	AI1843	hypothetical prote
3	39	63.9	82	2	S35770	T-cell receptor al
4	38	62.3	221	2	JC7307	endo-1,4-beta-xyla
5	38	62.3	394	2	G69230	endo-1,4-beta-gluc
6	37	60.7	71	2	E85976	probable galactosa
7	37	60.7	111	2	S57889	T cell receptor Er
8	37	60.7	245	2	AH3181	short chain dehydr
9	37	60.7	251	2	A65104	probable galactosa
10	37	60.7	572	2	H85731	Rhs element associ
11	37	60.7	913	2	AH0989	hypothetical ABC t
12	37	60.7	1398	2	H90698	RhsD core protein
13	37	60.7	1398	2	B85549	hypothetical prote
14	37	60.7	1400	2	E90886	RhsE core protein
15	37	60.7	1426	2	H64780	rhsD protein precu
16	36	59.0	72	2	AG2722	hypothetical prote
17	36	59.0	100	2	AG0447	probable phage-rel
18	36	59.0	357	2	D82337	UDP-N-acetylenolpy
19	35	57.4	116	1	QQECR3	hypothetical prote
20	35	57.4	363	1	WZBE44	gene 44 protein -
21	35	57.4	423	2	B75503	sporulation protei
22	35	57.4	448	2	B45438	myosin I beta, MMI
23	35	57.4	807	2	H75634	myosin-Ic - mouse
24	35	57.4	1028	2	S41749	myosin heavy chain
25	35	57.4	1028	2	S37146	myosin I heavy cha
26	35	57.4	1028	2	A59253	myosin I beta - hu
27	35	57.4	1095	2	T13964	probable histone d
28	34	55.7	130	2	S00827	T-cell receptor (R
29	34	55.7	215	2	T00938	hypothetical prote
30	34	55.7	300	2	S41171	transrepressor pro
31	34	55.7	335	2	D71275	hypothetical prote
32	34	55.7	354	2	E82850	fimbrial adhesin p
33	34	55.7	356	2	D82850	fimbrial adhesin p
34	34	55.7	422	2	E83083	probable two-compo
35	34	55.7	445	2	T06051	hypothetical prote
36	34	55.7	520	2	S72324	orotate reductase
37	34	55.7	646	2	B84852	probable receptor-
38	34	55.7	814	1	I39627	nicotine dehydroge
39	34	55.7	1145	2	B75625	hypothetical prote
40	33.5	54.9	350	2	H70929	probable PPE prote
41	33	54.1	23	2	S47194	T-cell receptor J-
42	33	54.1	261	2	A83196	probable transcrip
43	33	54.1	270	2	G70812	hypothetical prote
44	33	54.1	271	2	B96773	hypothetical prote
45	33	54.1	334	1	XYNHAL	site-specific DNA-

ALIGNMENTS

RESULT 1
A41912
T-cell receptor alpha chain (D10) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 17-Mar-1999
C;Accession: A41912
R;Hong, S.C.; Chelouche, A.; Lin, R.; Shaywitz, D.; Braunstein, N.S.; Glimcher, L.; J Cell 69, 999-1009, 1992
A;Title: An MHC interaction site maps to the amino-terminal half of the T cell recept
A;Reference number: A41912; MUID:92298397
A;Accession: A41912
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-63 <HON>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

Query Match 67.2%; Score 41; DB 2; Length 63;
Best Local Similarity 66.7%; Pred. No. 0.62;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 CLATGSARQLTF 12
           | ||||  :|||
Db      36 CAATGSFNKLTF 47
```

RESULT 2
AI1843
hypothetical protein alr0297 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AI1843
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI1843
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-384 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB77821.1; PID:g17135275; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr0297
C;Superfamily: hypothetical protein MJ1598

Query Match 65.6%; Score 40; DB 2; Length 384;
Best Local Similarity 63.6%; Pred. No. 5.4;

```
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLT 11
           ||:||:| :||
Db     122 CLSTGAAMELT 132


RESULT   3
S35770
T-cell receptor alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 23-Jul-1999
C;Accession: S35770
R;Wedderburn, L.R.
submitted to the EMBL Data Library, June 1993
A;Reference number: S35769
A;Accession: S35770
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-82 <WED>
A;Cross-references: EMBL:Z22966; NID:g312155; PIDN:CAA80539.1; PID:g312156
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

  Query Match             63.9%;  Score 39;  DB 2;  Length 82;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 GSARQLTF 12
           ||||||||
Db      60 GSARQLTF 67


RESULT   4
JC7307
endo-1,4-beta-xylanase (EC 3.2.1.8) - Penicillium sp.
N;Alternate names: endo-1,3-beta-xylanase; xylanase
C;Species: Penicillium sp.
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C;Accession: JC7307; PC7086
R;Kimura, T.; Ito, J.; Kawano, A.; Makino, T.; Kondo, H.; Karita, S.; Sakka, K.; Ohmiya,
Biosci. Biotechnol. Biochem. 64, 1230-1237, 2000
A;Title: Purification, characterization, and molecular cloning of acidophilic xylanase f
A;Reference number: JC7307
A;Accession: JC7307
A;Molecule type: DNA
A;Residues: 1-221 <KIM>
A;Cross-references: DDBJ:AB035540
A;Accession: PC7086
A;Molecule type: protein
A;Residues: 32-51 <KI2>
C;Genetics:
A;Gene: xynA
A;Introns: 89/2
C;Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology
C;Keywords: glycoprotein; glycosidase; hydrolase

  Query Match             62.3%;  Score 38;  DB 2;  Length 221;
  Best Local Similarity   70.0%;  Pred. No. 7.9;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
           :|||||  :||
Db      84 STGSARDITF 93


RESULT   5
G69230
endo-1,4-beta-glucanase related protein - Methanobacterium thermoautotrophicum (strain D
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: G69230
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: G69230
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-394 <MTH>
A;Cross-references: GB:AE000871; GB:AE000666; NID:g2622069; PIDN:AAB85473.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH977
A;Start codon: TTG

  Query Match             62.3%;  Score 38;  DB 2;  Length 394;
  Best Local Similarity   63.6%;  Pred. No. 14;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LATGSARQLTF 12
           ||||  |::||
Db     232 LATGETREVTF 242


RESULT   6
E85976
probable galactosamine-6-phosphate isomerase agaI_1 [imported] - Escherichia coli (st
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85976
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-71 <STO>
A;Cross-references: GB:AE005174; NID:g12517742; PIDN:AAG58273.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: agaI_1

  Query Match             60.7%;  Score 37;  DB 2;  Length 71;
  Best Local Similarity   58.3%;  Pred. No. 4.3;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           |||||:   ||:
Db      52 CLATGATPLLTY 63


RESULT   7
S57889
T cell receptor Er3 alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 23-Jul-1999
C;Accession: S57889
R;Giegerich, G.; Pette, M.; Meinl, E.; Epplen, J.T.; Wekerle, H.; Hinkkanen, A.
Eur. J. Immunol. 22, 753-758, 1992
A;Title: Diversity of T cell receptor alpha and beta chain genes expressed by human T
A;Reference number: S57869; MUID:92192091
A;Accession: S57889
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-111 <GIE>
A;Cross-references: EMBL:X57532; NID:g642982; PIDN:CAA40753.1; PID:g642983
C;Superfamily: immunoglobulin V region; immunoglobulin homology
```

C;Keywords: T-cell receptor

Query Match 60.7%; Score 37; DB 2; Length 111;
Best Local Similarity 50.0%; Pred. No. 6.6;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|: :|:| :|||
Db 88 CVCSGTASKLTF 99

RESULT 8
AH3181
short chain dehydrogenase Atu5181 [imported] - Agrobacterium tumefaciens (strain C58, Du
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AH3181
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AH3181
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-245 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL45870.1; PID:g17743613; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5181
A;Genome: plasmid
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 60.7%; Score 37; DB 2; Length 245;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LATGSARQLT 11
||:|:|||:|
Db 224 LASGAARQIT 233

RESULT 9
A65104
probable galactosamine-6-phosphate isomerase (EC 5.3.1.-) - Escherichia coli (strain K-1
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 16-Jul-1999
C;Accession: A65104
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65104
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-251 <BLAT>
A;Cross-references: GB:AE000395; GB:U00096; NID:g1789524; PIDN:AAC76175.1; PID:g1789530;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: agaI
C;Superfamily: glucosamine-6-phosphate isomerase
C;Keywords: intramolecular oxidoreductase; isomerase

Query Match 60.7%; Score 37; DB 2; Length 251;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|||||: ||:
Db 52 CLATGATPLLTY 63

RESULT 10
H85731
Rhs element associated protein Z2259 [imported] - Escherichia coli (strain O157:H7, s
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: H85731
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85731
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-572 <STO>
A;Cross-references: GB:AE005174; NID:g12515237; PIDN:AAG56316.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2259

Query Match 60.7%; Score 37; DB 2; Length 572;
Best Local Similarity 70.0%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ATGSARQLTF 12
|||| ||:|:
Db 389 ATGSTRQMTW 398

RESULT 11
AH0989
hypothetical ABC transporter ATP-binding protein STY4224 [imported] - Salmonella ente
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH0989
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; PMID:11677608
A;Accession: AH0989
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-913 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08044.1; PID:g16505025; GSPDB:GN00176
C;Genetics:
A;Gene: STY4224
C;Superfamily: Escherichia coli probable ABC transporter yhiH; ATP-binding cassette h

Query Match 60.7%; Score 37; DB 2; Length 913;
Best Local Similarity 88.9%; Pred. No. 48;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LATGSARQL 10
|||||||:||
Db 227 LATGSAQQL 235

RESULT 12
H90698
RhsD core protein with extension [imported] - Escherichia coli (strain O157:H7, subst
C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: H90698
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90698
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1398 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33983.1; PID:g13360018; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0560
C;Superfamily: rhsF protein

Query Match 60.7%; Score 37; DB 2; Length 1398;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ATGSARQLTF 12
|||| ||:|:
Db 531 ATGSTRQMTW 540

RESULT 13
B85549
hypothetical protein Z0651 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85549
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85549
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1398 <STO>
A;Cross-references: GB:AE005174; NID:g12513390; PIDN:AAG54854.1; GSPDB:GN00145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0651
C;Superfamily: rhsF protein

Query Match 60.7%; Score 37; DB 2; Length 1398;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ATGSARQLTF 12
|||| ||:|:
Db 531 ATGSTRQMTW 540

RESULT 14
E90886
RhsE core protein with extension [imported] - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: E90886
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90886
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1400 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35484.1; PID:g13361527; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2061
C;Superfamily: rhsF protein

Query Match 60.7%; Score 37; DB 2; Length 1400;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ATGSARQLTF 12
|||| ||:|:
Db 531 ATGSTRQMTW 540

RESULT 15
H64780
rhsD protein precursor - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Sep-1999
C;Accession: H64780; JS0625; B30092; I69401; S16026
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: H64780
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1426 <BLAT>
A;Cross-references: GB:AE000156; GB:U00096; NID:g1786705; PIDN:AAC73599.1; PID:g17867
A;Experimental source: strain K-12, substrain MG1655
R;Sadosky, A.B.; Gray, J.A.; Hill, C.W.
Nucleic Acids Res. 19, 7177-7183, 1991
A;Title: The RhsD-E subfamily of Escherichia coli K-12.
A;Reference number: JS0625; MUID:92115567
A;Accession: JS0625
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-378,'A',380-1166,'G',1168-1426 <SAD>
A;Cross-references: EMBL:X60999; NID:g42732; PID:g42733
A;Experimental source: strain K-12
R;Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
J. Bacteriol. 171, 636-642, 1989
A;Title: rhs gene family of Escherichia coli K-12.
A;Reference number: A91901; MUID:89123133
A;Accession: B30092
A;Molecule type: DNA
A;Residues: 1-100 <SA2>
A;Cross-references: GB:M21764; GB:J04224; NID:g147646; PIDN:AAA24542.1; PID:g147649
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A
J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhs
A;Reference number: I54935; MUID:90094253
A;Accession: I69401
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1232-1426 <RES>
A;Cross-references: GB:M29719; NID:g147644; PIDN:AAA24541.1; PID:g147645
C;Comment: the rhs core consist of two distinct parts: a large N-terminal core that i
C;Genetics:
A;Gene: rhsD
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1426/Product: rhsD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TMM>

Query Match 60.7%; Score 37; DB 2; Length 1426;
Best Local Similarity 70.0%; Pred. No. 72;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

10

```
Qy      3 ATGSARQLTF 12
          |||| ||:|:
Db    528 ATGSTRQMTW 537
```

Search completed: June 6, 2002, 06:16:24
Job time: 68 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:22:27 ; Search time 13.46 Seconds
(without alignments)
34.520 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	62.3	189	1	MOBA_PASMU	P57971 pasteurella
2	37	60.7	251	1	AGAI_ECOLI	P42912 escherichia
3	37	60.7	1426	1	RHSD_ECOLI	P16919 escherichia
4	36	59.0	210	1	GL34_ARATH	Q9flt3 arabidopsis
5	35	57.4	114	1	RT02_ASHGO	Q9hf53 ashbya goss
6	35	57.4	363	1	UL16_VZVD	P09293 varicella-z
7	35	57.4	1028	1	MY1C_HUMAN	O00159 homo sapien
8	35	57.4	1028	1	MY1C_MOUSE	Q9wti7 mus musculu
9	35	57.4	1149	1	HDA6_MOUSE	Q9z2v5 mus musculu
10	34	55.7	335	1	Y839_TREPA	O83811 treponema p
11	34	55.7	2282	1	ZAN_RABIT	P57999 oryctolagus
12	33	54.1	315	1	SX22_HUMAN	O15370 homo sapien
13	33	54.1	334	1	MTN3_NEILA	P24582 neisseria l
14	33	54.1	336	1	G3P_CORGL	Q01651 corynebacte
15	33	54.1	357	1	PMFE_PROMI	P53522 proteus mir
16	33	54.1	363	1	PGL1_COLLN	Q00446 colletotric
17	33	54.1	380	1	PGL2_PENOL	Q9y833 penicillium
18	33	54.1	458	1	RXRB_RAT	P49743 rattus norv
19	33	54.1	496	1	ERG1_YEAST	P32476 saccharomyc
20	33	54.1	520	1	RXRB_MOUSE	P28704 mus musculu
21	33	54.1	533	1	RXRB_HUMAN	P28702 homo sapien
22	33	54.1	673	1	FLID_TREMA	Q9kww7 treponema m
23	33	54.1	894	1	YHIH_ECOLI	P37624 escherichia
24	33	54.1	1505	1	PK3G_RAT	O70173 rattus norv
25	33	54.1	2813	1	VWF_CANFA	Q28295 canis famil
26	32	52.5	107	1	YB11_BACHD	Q9kdv0 bacillus ha
27	32	52.5	216	1	UPP_PORGI	Q9znf8 porphyromon
28	32	52.5	312	1	TRPF_CRYNE	P27710 cryptococcu
29	32	52.5	362	1	AROB_SALTY	P77980 salmonella
30	32	52.5	363	1	PGLA_ASPFL	P41749 aspergillus
31	32	52.5	363	1	PGLR_ASPOR	P35335 aspergillus
32	32	52.5	363	1	PGLR_ASPPA	P49575 aspergillus
33	32	52.5	366	1	CAS4_EPHMU	P18503 ephydatia m
34	32	52.5	496	1	E1BL_ADE05	P03243 human adeno
35	32	52.5	534	1	CN9A_MOUSE	O70628 mus musculu
36	32	52.5	556	1	RORA_HUMAN	P35398 homo sapien
37	32	52.5	564	1	BMRP_CANAL	P28873 candida alb
38	32	52.5	593	1	CN9A_HUMAN	O76083 homo sapien
39	32	52.5	610	1	YQ5C_CAEEL	Q09466 caenorhabdi
40	32	52.5	705	1	STT3_HUMAN	P46977 homo sapien
41	32	52.5	811	1	Y104_SYNY3	P54371 synechocyst
42	32	52.5	1156	1	GLH4_CAEEL	O76743 caenorhabdi
43	32	52.5	1873	1	CCAS_HUMAN	Q13698 homo sapien
44	32	52.5	2482	1	VWF_PIG	Q28833 sus scrofa
45	32	52.5	2700	1	ZAN_HUMAN	Q9y493 homo sapien

ALIGNMENTS

RESULT 1
MOBA_PASMU
ID MOBA_PASMU STANDARD; PRT; 189 AA.
AC P57971;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable molybdopterin-guanine dinucleotide biosynthesis protein A.
GN MOBA OR PM1799.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYDOPTERIN (MPT)
CC FORMING MOLYBDOPTERIN GUANINE DINUCLEOTIDE (MGD) (BY SIMILARITY).
CC -!- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MOBA FAMILY.
CC ----------

CC ----------
DR EMBL; AE006217; AAK03883.1; -.
DR HSSP; P32173; 1E5K.
KW Molybdenum cofactor biosynthesis; GTP-binding; Complete proteome.
SQ SEQUENCE 189 AA; 20922 MW; A467B28E1F42C2C7 CRC64;

Query Match 62.3%; Score 38; DB 1; Length 189;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 LATGSARQLTF 12
           ||||  | |||
Db     150 LATGERRMLTF 160
```

RESULT 2
AGAI_ECOLI
ID AGAI_ECOLI STANDARD; PRT; 251 AA.
AC P42912;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```
DE   Putative galactosamine-6-phosphate isomerase (EC 5.3.1.-)
DE   (Galactosamine-6-phosphate deaminase).
GN   AGAI OR B3141.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C;
RX   MEDLINE=20392444; PubMed=10931310;
RA   Brinkkoetter A., Kloess H., Alpert C.-A., Lengeler J.W.;
RT   "Pathways for the utilization of N-acetyl-galactosamine and
RT   galactosamine in Escherichia coli.";
RL   Mol. Microbiol. 37:125-135(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=97086503; PubMed=8932697;
RA   Reizer J., Ramseier T.M., Reizer A., Charbit A., Saier M.H. Jr.;
RT   "Novel phosphotransferase genes revealed by bacterial genome
RT   sequencing: a gene cluster encoding a putative N-acetylgalactosamine
RT   metabolic pathway in Escherichia coli.";
RL   Microbiology 142:231-250(1996).
CC   -!- CATALYTIC ACTIVITY: D-galactosamine 6-phosphate + H(2)O =
CC       D-tagatose 6-phosphate + NH(3).
CC   -!- PATHWAY: N-ACETYL GALACTOSAMINE UTILIZATION PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE
CC       ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF228498; AAF81093.1; -.
DR   EMBL; U18997; AAA57944.1; -.
DR   EMBL; AE000395; AAC76175.1; -.
DR   EcoGene; EG12772; agaI.
DR   InterPro; IPR000457; Glucosamine_iso.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   PROSITE; PS01161; GLC_GALNAC_ISOMERASE; 1.
KW   Isomerase; Complete proteome.
FT   ACT_SITE     86      86       BY SIMILARITY.
SQ   SEQUENCE   251 AA;  27724 MW;  823CB4CFBD82B8C6 CRC64;


  Query Match             60.7%;  Score 37;  DB 1;  Length 251;
  Best Local Similarity   58.3%;  Pred. No. 4.7;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           |||||:   ||:
Db      52 CLATGATPLLTY 63


RESULT   3
RHSD_ECOLI
ID   RHSD_ECOLI     STANDARD;      PRT;  1426 AA.
AC   P16919; P77232;
DT   01-AUG-1990 (Rel. 15, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   RhsD protein precursor.
GN   RHSD OR B0497.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92115567; PubMed=1766878;
RA   Sadosky A.B., Gray J.A., Hill C.W.;
RT   "The RhsD-E subfamily of Escherichia coli K-12.";
RL   Nucleic Acids Res. 19:7177-7183(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-176 FROM N.A.
RC   STRAIN=EC45;
RA   Wang Y.-D., Zhao S., Hill C.W.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-100 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=89123133; PubMed=2644231;
RA   Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;
RT   "rhs gene family of Escherichia coli K-12.";
RL   J. Bacteriol. 171:636-642(1989).
RN   [6]
RP   SEQUENCE OF 1232-1426 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=90094253; PubMed=2403547;
RA   Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,
RA   Vlazny D.A., Zhang J., Zhao S., Hill C.W.;
RT   "Structure of the rhsA locus from Escherichia coli K-12 and
RT   comparison of rhsA with other members of the rhs multigene family.";
RL   J. Bacteriol. 172:446-456(1990).
RN   [7]
RP   REVIEW.
RX   MEDLINE=95020608; PubMed=7934896;
RA   Hill C.W., Sandt C.H., Vlazny D.A.;
RT   "Rhs elements of Escherichia coli: a family of genetic composites
RT   each encoding a large mosaic protein.";
RL   Mol. Microbiol. 12:865-871(1994).
CC   -!- FUNCTION: RHS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY
CC       PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.
CC   -!- DOMAIN: EACH RHS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa
CC       AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE RHS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE000156; AAC73599.1; -.
DR   EMBL; U82664; AAB40251.1; -.
DR   EMBL; AF044500; AAC32467.1; -.
DR   EMBL; X60999; CAA43314.1; -.
DR   EMBL; M21764; AAA24542.1; -.
DR   PIR; B30092; B30092.
DR   PIR; JS0625; JS0625.
DR   PIR; S16026; S16026.
DR   EcoGene; EG10849; rhsD.
DR   InterPro; IPR001826; RHS.
DR   PRINTS; PR00394; RHSPROTEIN.
KW   Signal; Multigene family; Transmembrane; Repeat; Complete proteome.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27   1426       RHSD PROTEIN.
FT   TRANSMEM    28     55       POTENTIAL.
FT   CONFLICT   379    379       G -> A (IN REF. 1).
FT   CONFLICT  1167   1167       A -> G (IN REF. 1).
SQ   SEQUENCE  1426 AA;  159724 MW;  122D697AA449BC2F CRC64;


  Query Match             60.7%;  Score 37;  DB 1;  Length 1426;
  Best Local Similarity   70.0%;  Pred. No. 27;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 ATGSARQLTF 12
          |||| ||:|:
Db    528 ATGSTRQMTW 537


RESULT   4
GL34_ARATH
ID   GL34_ARATH     STANDARD;      PRT;   210 AA.
AC   Q9FLT3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative germin-like protein subfamily 3 member 4 precursor.
GN   AT5G61750 OR MAC9.6 OR MAC9_50.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016721; PubMed=11130714;
RA   Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA   Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA   Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA   Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA   Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA   Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA   Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA   Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA   Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA   Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA   Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA   Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA   Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA   Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA   Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA   Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA   Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA   Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA   van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA   Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA   Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA   Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA   Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;
RT   "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:823-826(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=98290546; PubMed=9628582;
RA   Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT   Sequence features of the regions of 1,456,315 bp covered by nineteen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:41-54(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
CC       OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
CC   -!- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
CC       HEXAMER) (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Apoplast (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AB010069; BAB10075.1; -.
DR   InterPro; IPR001929; Germin.
DR   Pfam; PF01072; Germin; 1.
DR   PRINTS; PR00325; GERMIN.
DR   PROSITE; PS00725; GERMIN; 1.
KW   Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KW   Multigene family; Hypothetical protein.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19    210       PUTATIVE GERMIN-LIKE PROTEIN
FT                               SUBFAMILY 3 MEMBER 4.
FT   METAL      106    106       MANGANESE (BY SIMILARITY).
FT   METAL      108    108       MANGANESE (BY SIMILARITY).
FT   METAL      113    113       MANGANESE (BY SIMILARITY).
FT   METAL      152    152       MANGANESE (BY SIMILARITY).
FT   DISULFID    27     44       BY SIMILARITY.
FT   CARBOHYD    73     73       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE  210 AA;  22890 MW;  F80695370A6F5284 CRC64;


  Query Match             59.0%;  Score 36;  DB 1;  Length 210;
  Best Local Similarity   63.6%;  Pred. No. 6.3;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CLATGSARQLT 11
          ||||||  ::|
Db    195 CLATGSGCRVT 205


RESULT   5
RT02_ASHGO
ID   RT02_ASHGO     STANDARD;      PRT;   114 AA.
AC   Q9HF53;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mitochondrial 40S ribosomal protein MRP2.
GN   MRP2.
OS   Ashbya gossypii (Eremothecium gossypii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Eremotheciaceae; Eremothecium.
OX   NCBI_TaxID=33169;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendland J., Dietrich F.S., Philippsen P.;
RT   "Hyphal morphogenesis is controlled by multiple Rho-protein modules in
RT   the filamentous ascomycete Ashbya gossypii.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF210628; AAG41250.1; -.
DR   InterPro; IPR001209; Ribosomal_S14.
DR   Pfam; PF00253; Ribosomal_S14; 1.
DR   PROSITE; PS00527; RIBOSOMAL_S14; 1.
KW   Ribosomal protein; Mitochondrion.
SQ   SEQUENCE   114 AA;  13017 MW;  BD15BF5A2107F02F CRC64;


  Query Match             57.4%;  Score 35;  DB 1;  Length 114;
  Best Local Similarity   60.0%;  Pred. No. 5.4;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQL 10
           |:||| || :
Db      77 CIATGHARSV 86


RESULT   6
UL16_VZVD
ID   UL16_VZVD      STANDARD;      PRT;   363 AA.
AC   P09293;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Gene 44 protein.
GN   44.
OS   Varicella-zoster virus (strain Dumas) (VZV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10338;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86306657; PubMed=3018124;
RA   Davison A.J., Scott J.E.;
RT   "The complete DNA sequence of varicella-zoster virus.";
RL   J. Gen. Virol. 67:1759-1816(1986).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,
CC       HSV-6 ORF11R, EHV-1 46, HCMV UL94, EBV BGLF2, HSV 33, AND VZV 44.
CC   ----------------------------------------------------------------------
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04370; CAA27927.1; -.
DR   PIR; I27341; WZBE44.
DR   InterPro; IPR004286; UL16_UL94.
DR   Pfam; PF03044; UL16_UL94; 1.
SQ   SEQUENCE   363 AA;  40244 MW;  5F78F627E6261B87 CRC64;


  Query Match             57.4%;  Score 35;  DB 1;  Length 363;
  Best Local Similarity   77.8%;  Pred. No. 17;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLT 11
           |||:||:||
Db      12 ATGAARKLT 20


RESULT   7
MY1C_HUMAN
ID   MY1C_HUMAN     STANDARD;      PRT;  1028 AA.
AC   O00159;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Ic (Myosin I beta) (MMI-beta) (MMIb).
GN   MYO1C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=97237053; PubMed=9119401;
RA   Crozet F., Amraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA   Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA   Weil D., Pujol R., Petit C.;
RT   "Cloning of the genes encoding two murine and human cochlear
RT   unconventional type I myosins.";
RL   Genomics 40:332-341(1997).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   ----------------------------------------------------------------------
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X98507; CAA67131.1; -.
DR   HSSP; P08799; 1LVK.
DR   MIM; 606538; -.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 2.
KW   Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW   Multigene family.
FT   DOMAIN        1    683       MYOSIN HEAD-LIKE.
FT   DOMAIN      699    722       IQ 1.
FT   DOMAIN      723    751       IQ 2.
FT   NP_BIND     105    112       ATP (POTENTIAL).
SQ   SEQUENCE   1028 AA;  118037 MW;  0E9C3680527F85C6 CRC64;


  Query Match             57.4%;  Score 35;  DB 1;  Length 1028;
  Best Local Similarity   58.3%;  Pred. No. 49;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           ||  | |  |||
Db     463 CLRPGEATDLTF 474


RESULT   8
MY1C_MOUSE
ID   MY1C_MOUSE     STANDARD;      PRT;  1028 AA.
```

```
AC   Q9WTI7; O08834; O08571; Q9QW54;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Ic (Myosin I beta) (MMIb).
GN   MYO1C.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC   STRAIN=DBA/2J; TISSUE=Brain;
RX   MEDLINE=97325785; PubMed=9182797;
RA   Hamilton B.A., Smith D.J., Mueller K.L., Kerrebrock A.W.,
RA   Bronson R.T., van Berkel V., Daly M.J., Kruglyak L., Reeve M.P.,
RA   Nemhauser J.L., Hawkins T.L., Rubin E.M., Lander E.S.;
RT   "The vibrator mutation causes neurodegeneration via reduced expression
RT   of PITP alpha: positional complementation cloning and extragenic
RT   suppression.";
RL   Neuron 18:711-722(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE OF 1-807 FROM N.A. (ISOFORM B).
RC   STRAIN=BALB/C; TISSUE=Cochlea;
RX   MEDLINE=97237053; PubMed=9119401;
RA   Crozet F., Amraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA   Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA   Weil D., Pujol R., Petit C.;
RT   "Cloning of the genes encoding two murine and human cochlear
RT   unconventional type I myosins.";
RL   Genomics 40:332-341(1997).
RN   [4]
RP   SEQUENCE OF 146-593 FROM N.A.
RX   MEDLINE=93194946; PubMed=8449986;
RA   Sherr E.H., Joyce M.P., Greene L.A.;
RT   "Mammalian myosin I alpha, I beta, and I gamma: new widely expressed
RT   genes of the myosin I family.";
RL   J. Cell Biol. 120:1405-1416(1993).
CC   -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC       ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC       THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC       COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC       SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U96723; AAC53264.1; -.
DR   EMBL; AK004743; BAB23524.1; -.
DR   EMBL; U96726; AAC60758.1; -.
DR   EMBL; X99638; CAA67956.1; -.
DR   HSSP; P08799; 1LVK.
DR   MGD; MGI:106612; Myo1c.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 3.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 2.
KW   Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW   Multigene family; Alternative splicing.
FT   DOMAIN        1    683       MYOSIN HEAD-LIKE.
FT   DOMAIN      699    722       IQ 1.
FT   DOMAIN      723    751       IQ 2.
FT   NP_BIND     105    112       ATP (POTENTIAL).
FT   VARSPLIC    897   1028       GYKPRPRQLLLTPSAVVIVEDAKVKQRIDYANLTGISVSSL
FT                                SDSLFVLHVQREDNKQKGDVVLQSDHVIETLTKTALSADRV
FT                                NNININQGSITFAGGPGRDGIIDFTSGSELLITKAKNGHLA
FT                                VVAPRLNSR -> VTSLAPGSCCSRPVLWSLWRMLKSSREL
FT                                IMPT (IN ISOFORM B).
FT   CONFLICT     34     35       RR -> GG (IN REF. 3).
FT   CONFLICT     62     64       SRQ -> RRK (IN REF. 3).
FT   CONFLICT    216    216       T -> A (IN REF. 4).
FT   CONFLICT    237    237       C -> F (IN REF. 4).
FT   CONFLICT    251    252       VM -> LL (IN REF. 4).
FT   CONFLICT    353    353       R -> A (IN REF. 4).
FT   CONFLICT    366    368       LAS -> VPA (IN REF. 4).
FT   CONFLICT    411    411       Q -> R (IN REF. 4).
FT   CONFLICT    482    484       VKP -> IKH (IN REF. 4).
FT   CONFLICT    543    543       M -> T (IN REF. 4).
FT   CONFLICT    572    572       S -> G (IN REF. 4).
FT   CONFLICT    700    700       Q -> R (IN REF. 3).
FT   CONFLICT    786    786       R -> G (IN REF. 3).
FT   CONFLICT    807    807       E -> D (IN REF. 3).
SQ   SEQUENCE   1028 AA;  118155 MW;  2CF5535D816884DF CRC64;
```

Query Match 57.4%; Score 35; DB 1; Length 1028;
Best Local Similarity 58.3%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 CLATGSARQLTF 12
           ||  | |  |||
Db     463 CLRPGEATDLTF 474
```

```
RESULT   9
HDA6_MOUSE
ID   HDA6_MOUSE     STANDARD;      PRT;  1149 AA.
AC   Q9Z2V5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone deacetylase 6 (HD6) (Histone deacetylase mHDA2).
GN   HDAC6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Fetal;
RX   MEDLINE=99107904; PubMed=9891014;
RA   Verdel A., Khochbin S.;
RT   "Identification of a new family of higher eukaryotic histone
RT   deacetylases. Coordinate expression of differentiation-dependent
RT   chromatin modifiers.";
RL   J. Biol. Chem. 274:2440-2445(1999).
CC   -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC       THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC       HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC       REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC       FAMILY. HD SUBFAMILY 2.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF006603; AAD09835.2; -.
DR   MGD; MGI:1333752; Hdac6.
DR   InterPro; IPR000286; His_deacetylse.
DR   InterPro; IPR001607; zf-UBP.
DR   Pfam; PF00850; Hist_deacetyl; 2.
DR   Pfam; PF02148; zf-UBP; 1.
DR   PRINTS; PR01270; HDASUPER.
DR   SMART; SM00290; ZnF_UBP; 1.
KW   Hydrolase; Nuclear protein; Repeat.
FT   DOMAIN       87    403       HISTONE DEACETYLASE 1.
FT   DOMAIN      481    799       HISTONE DEACETYLASE 2.
FT   DOMAIN      455    460       POLY-GLU.
SQ   SEQUENCE   1149 AA;  125703 MW;  2B98CDB228CE0D1D CRC64;


  Query Match             57.4%;  Score 35;  DB 1;  Length 1149;
  Best Local Similarity   70.0%;  Pred. No. 55;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQL 10
           ||||||  :|
Db     184 CLATGSVLRL 193


RESULT  10
Y839_TREPA
ID   Y839_TREPA     STANDARD;      PRT;   335 AA.
AC   O83811;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical lipoprotein TP0839 precursor.
GN   TP0839.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001254; AAC65809.1; -.
DR   TIGR; TP0839; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Membrane; Lipoprotein; Signal;
KW   Complete proteome.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    335       HYPOTHETICAL LIPOPROTEIN TP0839.
FT   LIPID        22     22       N-ACYL DIGLYCERIDE (POTENTIAL).
SQ   SEQUENCE   335 AA;  37364 MW;  2F4C5FE154DAB869 CRC64;


  Query Match             55.7%;  Score 34;  DB 1;  Length 335;
  Best Local Similarity   60.0%;  Pred. No. 25;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQL 10
           ||:|| : ||
Db     311 CLSTGGSAQL 320


RESULT  11
ZAN_RABIT
ID   ZAN_RABIT      STANDARD;      PRT;  2282 AA.
AC   P57999;
DT   16-OCT-2001 (Rel. 40, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Zonadhesin (Fragment).
GN   ZAN.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Lea I.A., Sivashanmugam P., Richardson R.T., O'Rand M.G.;
RT   "Sequence of rabbit zonadhesin.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC       OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC       SIGNALING (BY SIMILARITY).
CC   -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC       APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC   -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC       ZONA PELLUCIDA.
CC   -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC       THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC       SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC   -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC       OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC   -!- SIMILARITY: CONTAINS AT LEAST 2 MAM DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; AF244982; AAF63342.2; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR003328; TILa.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00629; MAM; 2.
DR   Pfam; PF01826; TIL; 5.
DR   Pfam; PF02345; TILa; 5.
DR   Pfam; PF00094; vwd; 4.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00137; MAM; 1.
DR   SMART; SM00214; VWC; 5.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS00740; MAM_1; FALSE_NEG.
DR   PROSITE; PS50060; MAM_2; 2.
KW   Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW   Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1   2235       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2236   2256       POTENTIAL.
FT   DOMAIN     2257   2282       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       <1    147       MAM 1.
FT   DOMAIN      150    315       MAM 2.
FT   DOMAIN      315    498       26 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT                                (MUCIN-LIKE DOMAIN).
FT   DOMAIN      499    610       VWFD 1 (PARTIAL).
FT   DOMAIN      611    995       VWFD 2.
FT   DOMAIN      996   1384       VWFD 3.
FT   DOMAIN     1385   1787       VWFD 4.
FT   DOMAIN     1788   2189       VWFD 5.
FT   DOMAIN     2185   2221       EGF-LIKE.
FT   DISULFID   2189   2200       BY SIMILARITY.
FT   DISULFID   2194   2209       BY SIMILARITY.
FT   DISULFID   2211   2220       BY SIMILARITY.
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    541    541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    569    569       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1141   1141       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1259   1259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1270   1270       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1355   1355       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1467   1467       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1483   1483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1662   1662       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1997   1997       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2178   2178       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2282 AA;  248290 MW;  380FA81093454892 CRC64;


  Query Match           55.7%;  Score 34;  DB 1;  Length 2282;
  Best Local Similarity  50.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           |  || :| |:|
Db    1386 CQVTGDSRYLSF 1397


RESULT  12
SX22_HUMAN
ID   SX22_HUMAN     STANDARD;      PRT;   315 AA.
AC   O15370; Q9NUD4;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   SOX-22 protein.
GN   SOX22.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=97358581; PubMed=9215677;
RA   Jay P., Sahly I., Goze C., Taviaux S., Poulat F., Couly G.,
RA   Abitbol M., Berta P.;
RT   "SOX22 is a new member of the SOX gene family, mainly expressed in
RT   human nervous tissue.";
RL   Hum. Mol. Genet. 6:1069-1077(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaslaiho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN THE CNS.
CC       ALSO EXPRESSED IN FETAL BRAIN AND KIDNEY AND ADULT HEART,
CC       PANCREAS, TESTIS AND OVARY. OTHER TISSUES WERE ONLY WEAKLY
CC       POSITIVE.
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; U35612; AAB69627.1; -.
DR   EMBL; AL034548; CAB81632.1; -.
DR   HSSP; P48436; 1SX9.
DR   MIM; 601947; -.
DR   InterPro; IPR000910; HMG_12_box.
DR   Pfam; PF00505; HMG_box; 1.
DR   SMART; SM00398; HMG; 1.
KW   DNA-binding; Nuclear protein.
FT   DOMAIN       15     22       POLY-PRO.
FT   DNA_BIND     40    108       HMG BOX.
FT   DOMAIN      211    217       POLY-ALA.
FT   DOMAIN      223    250       ASP/GLU-RICH (ACIDIC).
```

```
FT   DOMAIN       223    233       POLY-GLU.
FT   DOMAIN       234    237       POLY-ALA.
FT   CONFLICT     135    201       GSSPGRSCLAAGAAEQREGLWGAGRRRPRTTMKTTTRSCWK
FT                                 CAWSRPRGGSCGGWSRRDGPLGDKRS -> RLKPGPQLPGR
FT                                 GGRRAAGGPLGGGAAAPEDDDEDDDEELLEVRLVETPGREL
FT                                 WRMVPAGRAARGQAE (IN REF. 2).
SQ   SEQUENCE   315 AA;  34301 MW;  DE7DD00E6660B6DC CRC64;


  Query Match             54.1%;  Score 33;  DB 1;  Length 315;
  Best Local Similarity   66.7%;  Pred. No. 37;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           ||| |:| |
Db     142 CLAAGAAEQ 150


RESULT  13
MTN3_NEILA
ID   MTN3_NEILA     STANDARD;      PRT;   334 AA.
AC   P24582;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase NlaIII (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase NlaIII) (M.NlaIII).
GN   NLAIIIM.
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=486;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23970 / NRCC 2118;
RX   MEDLINE=91117164; PubMed=2277628;
RA   Labbe D., Hoeltke H.J., Lau P.C.K.;
RT   "Cloning and characterization of two tandemly arranged DNA
RT   methyltransferase genes of Neisseria lactamica: an adenine-specific
RT   M.NlaIII and a cytosine-type methylase.";
RL   Mol. Gen. Genet. 224:101-110(1990).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC       CATG, CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS AND
CC       PROTECTS THE DNA FROM CLEAVAGE BY THE NLAIII ENDONUCLEASE.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-57 IS THE INITIATOR.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54485; CAA38356.1; -.
DR   PIR; S12036; XYNHAL.
DR   REBASE; 3468; M.NlaIII.
DR   InterPro; IPR002294; D12N6_mtfrase.
DR   InterPro; IPR002052; N6_Mtase.
DR   Pfam; PF02086; MethyltransfD12; 1.
DR   PRINTS; PR00505; D12N6MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; Restriction system.
SQ   SEQUENCE   334 AA;  38382 MW;  8BCF1A1A35F02E89 CRC64;


  Query Match             54.1%;  Score 33;  DB 1;  Length 334;
  Best Local Similarity   66.7%;  Pred. No. 40;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           || :|| ||
Db     100 CLGSGSERQ 108


RESULT  14
G3P_CORGL
ID   G3P_CORGL      STANDARD;      PRT;   336 AA.
AC   Q01651;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
GN   GAP.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC   Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13059 / AS019;
RX   MEDLINE=93015645; PubMed=1400158;
RA   Eikmanns B.J.;
RT   "Identification, sequence analysis, and expression of a
RT   Corynebacterium glutamicum gene cluster encoding the three glycolytic
RT   enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate
RT   kinase, and triosephosphate isomerase.";
RL   J. Bacteriol. 174:6076-6086(1992).
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC       NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC   -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X59403; CAA42045.1; -.
DR   PIR; S23910; S23910.
DR   PIR; A43260; A43260.
DR   HSSP; P00362; 1GD1.
DR   InterPro; IPR000173; GAP_DH.
DR   Pfam; PF00044; gpdh; 1.
DR   Pfam; PF02800; gpdh_C; 1.
DR   PRINTS; PR00078; G3PDHDRGNASE.
DR   PROSITE; PS00071; GAPDH; 1.
KW   Glycolysis; Oxidoreductase; NAD.
FT   BINDING     153    153       GLYCERALDEHYDE 3-PHOSPHATE.
FT   ACT_SITE    180    180       ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ   SEQUENCE   336 AA;  36199 MW;  5514A0A0CF078219 CRC64;


  Query Match             54.1%;  Score 33;  DB 1;  Length 336;
  Best Local Similarity   77.8%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TGSARQLTF 12
           |||| |||
Db     239 TGSATDLTF 247


RESULT  15
PMFE_PROMI
ID   PMFE_PROMI     STANDARD;      PRT;   357 AA.
AC   P53522;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
```

```
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Putative minor fimbrial subunit pmfE precursor.
GN   PMFE.
OS   Proteus mirabilis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=584;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HI4320;
RX   MEDLINE=95047519; PubMed=7959033;
RA   Massad G., Mobley H.L.T.;
RT   "Genetic organization and complete sequence of the Proteus mirabilis
RT   pmf fimbrial operon.";
RL   Gene 150:101-104(1994).
CC   -!- SUBCELLULAR LOCATION: Fimbria.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35428; CAA84592.1; -.
KW   Fimbria; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    357       PUTATIVE MINOR FIMBRIAL SUBUNIT PMFE.
SQ   SEQUENCE   357 AA;  38876 MW;  59AD7E566D4899AA CRC64;


  Query Match             54.1%;. Score 33;  DB 1;  Length 357;
  Best Local Similarity   54.5%;  Pred. No. 42;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LATGSARQLTF 12
           :::||: |||| 
Db     154 ISSGSSGQLTF 164


Search completed: June  6, 2002, 06:22:29
Job time: 363 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:22:08 ; Search time 41.63 Seconds
(without alignments)
49.866 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	62.3	221	3	Q9UUQ2	Q9uuq2 penicillium
2	38	62.3	394	17	O27058	O27058 methanother
3	37	60.7	1252	11	Q9EQS9	Q9eqs9 mus musculu
4	37	60.7	1252	11	Q9JLI1	Q9jli1 mus musculu
5	37	60.7	1253	11	Q9EQS8	Q9eqs8 mus musculu
6	37	60.7	1409	2	O52661	O52661 escherichia
7	37	60.7	1426	2	Q93V17	Q93v17 escherichia
8	36	59.0	357	16	Q9KV40	Q9kv40 vibrio chol
9	35.5	58.2	1168	16	Q99YQ8	Q99yq8 streptococc
10	35	57.4	138	17	Q9HHZ4	Q9hhz4 halobacteri
11	35	57.4	309	5	O97470	O97470 dictyosteli
12	35	57.4	423	16	Q9RWU3	Q9rwu3 deinococcus
13	35	57.4	641	4	O95186	O95186 homo sapien
14	35	57.4	660	3	Q9C1M9	Q9c1m9 aspergillus
15	35	57.4	1028	6	Q27966	Q27966 bos taurus
16	35	57.4	1028	6	Q28138	Q28138 bos taurus
17	35	57.4	1028	11	Q63355	Q63355 rattus norv
18	35	57.4	1044	11	Q9ERB6	Q9erb6 mus musculu
19	34	55.7	84	2	Q93AG7	Q93ag7 uncultured
20	34	55.7	91	2	O24790	O24790 lactobacill
21	34	55.7	91	2	Q48499	Q48499 lactobacill
22	34	55.7	208	2	Q9WWM0	Q9wwm0 synechococc
23	34	55.7	266	10	Q9AWX0	Q9awx0 oryza sativ
24	34	55.7	300	12	Q89582	Q89582 bovine herp
25	34	55.7	347	5	Q9W1A3	Q9w1a3 drosophila
26	34	55.7	354	16	Q9PH68	Q9ph68 xylella fas
27	34	55.7	356	16	Q9PH69	Q9ph69 xylella fas
28	34	55.7	396	2	Q9ZF15	Q9zf15 streptomyce
29	34	55.7	422	16	Q9HVS7	Q9hvs7 pseudomonas
30	34	55.7	445	10	Q9SVK1	Q9svk1 arabidopsis
31	34	55.7	480	10	Q9M5G4	Q9m5g4 euphorbia e
32	34	55.7	501	10	Q9AYA6	Q9aya6 oryza sativ
33	34	55.7	520	3	Q12610	Q12610 emericella
34	34	55.7	646	10	Q9SLD0	Q9sld0 arabidopsis
35	34	55.7	749	10	Q9LHT9	Q9lht9 oryza sativ
36	34	55.7	814	2	Q59129	Q59129 arthrobacte
37	34	55.7	1145	16	Q9RZQ5	Q9rzq5 deinococcus
38	34	55.7	1849	2	Q9S4K2	Q9s4k2 lactobacill
39	33.5	54.9	350	16	O53940	O53940 mycobacteri
40	33	54.1	77	3	Q96WD2	Q96wd2 pseudogibel
41	33	54.1	220	3	Q9HDX5	Q9hdx5 schizosacch
42	33	54.1	238	10	Q9LDX8	Q9ldx8 oryza sativ
43	33	54.1	244	2	Q9X4U4	Q9x4u4 streptococc
44	33	54.1	261	16	Q9HY27	Q9hy27 pseudomonas
45	33	54.1	270	16	O53851	O53851 mycobacteri

ALIGNMENTS

RESULT 1
Q9UUQ2
ID Q9UUQ2 PRELIMINARY; PRT; 221 AA.
AC Q9UUQ2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE XYLANASE A (EC 3.2.1.8).
GN XYNA.
OS Penicillium sp. 40.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=110511;
RN [1]
RP SEQUENCE FROM N.A.
RA Kimura T., Sakka K., Ohmiya K.;
RT "Acidophilic xylanase A from Penicillium sp.40.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB035540; BAA88421.1; -.
DR HSSP; P36217; 1XYO.
DR InterPro; IPR001137; Glyco_hydro_11.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PR00911; GLHYDRLASE11.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 221 AA; 24171 MW; 98BDD5BCBEC860AD CRC64;

Query Match 62.3%; Score 38; DB 3; Length 221;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy       3 ATGSARQLTF 12
           :|||||  :||
Db      84 STGSARDITF 93
```

RESULT 2
O27058

```
ID   O27058      PRELIMINARY;      PRT;   394 AA.
AC   O27058;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   ENDO-1,4-BETA-GLUCANASE RELATED PROTEIN.
GN   MTH977.
OS   Methanothermobacter thermautotrophicus.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000871; AAB85473.1; -.
KW   Complete proteome.
SQ   SEQUENCE   394 AA;  42683 MW;  B4DD13CF7CAE9B8B CRC64;
```

```
  Query Match             62.3%;  Score 38;  DB 17;  Length 394;
  Best Local Similarity   63.6%;  Pred. No. 33;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LATGSARQLTF 12
           ||||  |::||
Db     232 LATGETREVTF 242
```

```
RESULT   3
Q9EQS9
ID   Q9EQS9      PRELIMINARY;      PRT;  1252 AA.
AC   Q9EQS9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DDM36.
GN   NOPE OR DDM36.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/CBA;
RA   Murakami H., Nakamata T., Nakayama T., Yamamoto H., Hosaka T.,
RA   Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,
RA   Toguchida J.;
RT   "Up-regulation of a ras effector and down-regulation of a cell
RT   adhesion molecule are associated with transformation of osteoblasts.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB052620; BAB19278.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1858497; Nope.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 5.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1252 AA;  134764 MW;  BA5292393483AB73 CRC64;
```

```
  Query Match             60.7%;  Score 37;  DB 11;  Length 1252;
  Best Local Similarity   77.8%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |:|| ||||
Db     211 CVATNSARQ 219
```

```
RESULT   4
Q9JLI1
ID   Q9JLI1      PRELIMINARY;      PRT;  1252 AA.
AC   Q9JLI1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEIGHBOR OF PUNC E11 PROTEIN.
GN   NOPE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB;
RX   MEDLINE=20175427; PubMed=10708514;
RA   Salbaum J.M., Kappen C.;
RT   "Cloning and expression of Nope, a new mouse gene of the
RT   immunoglobulin superfamily related to guidance receptors.";
RL   Genomics 64:15-23(2000).
DR   EMBL; AF176694; AAF65930.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1858497; Nope.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 5.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1252 AA;  134759 MW;  11948773277B76B2 CRC64;
```

```
  Query Match             60.7%;  Score 37;  DB 11;  Length 1252;
  Best Local Similarity   77.8%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |:|| ||||
Db     211 CVATNSARQ 219
```

```
RESULT   5
Q9EQS8
ID   Q9EQS8      PRELIMINARY;      PRT;  1253 AA.
AC   Q9EQS8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DDM36E.
```

```
GN   NOPE OR DDM36E.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/CBA;
RA   Murakami H., Nakamata T., Nakayama T., Yamamoto H., Hosaka T.,
RA   Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,
RA   Toguchida J.;
RT   "Up-regulation of a ras effector and down-regulation of a cell
RT   adhesion molecule are associated with transformation of osteoblasts.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB052621; BAB19279.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1858497; Nope.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 5.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1253 AA;  134894 MW;  2D55B254A42B9AEE CRC64;

  Query Match             60.7%;  Score 37;  DB 11;  Length 1253;
  Best Local Similarity   77.8%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
            |:|| ||||
Db     211 CVATNSARQ 219


RESULT   6
O52661
ID   O52661      PRELIMINARY;      PRT;  1409 AA.
AC   O52661;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CORE PROTEIN.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EC50;
RX   MEDLINE=98361897; PubMed=9696756;
RA   Wang Y.D., Zhao S., Hill C.W.;
RT   "Rhs elements comprise three subfamilies which diverged prior to
RT   acquisition by Escherichia coli.";
RL   J. Bacteriol. 180:4102-4110(1998).
DR   EMBL; AF044499; AAC32464.1; -.
DR   InterPro; IPR001826; RHS.
DR   PRINTS; PR00394; RHSPROTEIN.
SQ   SEQUENCE   1409 AA;  158612 MW;  2D5D82E8A11BDEA8 CRC64;

  Query Match             60.7%;  Score 37;  DB 2;  Length 1409;
  Best Local Similarity   70.0%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
            |||| ||:|:
Db     531 ATGSTRQMTW 540


RESULT   7
Q93V17
ID   Q93V17      PRELIMINARY;      PRT;  1426 AA.
AC   Q93V17;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CORE PROTEIN.
GN   RHSD.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K-12;
RX   MEDLINE=89123133; PubMed=2644231;
RA   Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;
RT   "rhs gene family of Escherichia coli K-12.";
RL   J. Bacteriol. 171:636-642(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K-12;
RX   MEDLINE=92115567; PubMed=1766878;
RA   Sadosky A.B., Gray J.A., Hill C.W.;
RT   "The RhsD-E subfamily of Escherichia coli K-12.";
RL   Nucleic Acids Res. 19:7177-7183(1991).
DR   EMBL; L19084; AAA24544.1; -.
SQ   SEQUENCE   1426 AA;  159724 MW;  4F6D7084470FBC2D CRC64;

  Query Match             60.7%;  Score 37;  DB 2;  Length 1426;
  Best Local Similarity   70.0%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
            |||| ||:|:
Db     528 ATGSTRQMTW 537


RESULT   8
Q9KV40
ID   Q9KV40      PRELIMINARY;      PRT;  357 AA.
AC   Q9KV40;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE REDUCTASE.
GN   VC0318.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004120; AAF93491.1; -.
DR   HSSP; P08373; 2MBR.
```

```
DR   TIGR; VC0318; -.
DR   InterPro; IPR003170; MurB.
DR   Pfam; PF02215; MurB; 1.
KW   Complete proteome.
SQ   SEQUENCE   357 AA;  39351 MW;  C5DC7E1EEE834E0C CRC64;


  Query Match             59.0%;  Score 36;  DB 16;  Length 357;
  Best Local Similarity   54.5%;  Pred. No. 74;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLT 11
           || ||: ::||
Db     154 CLETGTVKRLT 164


RESULT   9
Q99YQ8
ID   Q99YQ8      PRELIMINARY;      PRT;  1168 AA.
AC   Q99YQ8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE BETA-GALACTOSIDASE.
GN   SPY1586.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR   EMBL; AE006590; AAK34368.1; -.
DR   InterPro; IPR001649; Glyco_hydro_2.
DR   Pfam; PF00703; Glyco_hydro_2; 1.
DR   PRINTS; PR00132; GLHYDRLASE2.
KW   Complete proteome.
SQ   SEQUENCE   1168 AA;  131946 MW;  F323127F2449CEA3 CRC64;


  Query Match             58.2%;  Score 35.5;  DB 16;  Length 1168;
  Best Local Similarity   75.0%;  Pred. No. 3e+02;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 CLATGSARQLTF 12
           || ||  |||||
Db     318 CLETG-FRQLTF 328


RESULT  10
Q9HHZ4
ID   Q9HHZ4      PRELIMINARY;      PRT;   138 AA.
AC   Q9HHZ4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   VNG6157H.
GN   VNG6157H.
OS   Halobacterium sp. (strain NRC-1).
OG   Plasmid pNRC200.
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005152; AAG20828.1; -.
KW   Plasmid; Complete proteome.
SQ   SEQUENCE   138 AA;  14950 MW;  3E347A1BB44F2DAC CRC64;


  Query Match             57.4%;  Score 35;  DB 17;  Length 138;
  Best Local Similarity   54.5%;  Pred. No. 44;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLT 11
           ||: | || :|
Db      36 CLSQGEARSIT 46


RESULT  11
O97470
ID   O97470      PRELIMINARY;      PRT;   309 AA.
AC   O97470;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   ADP/ATP TRANSLOCASE.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX2;
RX   MEDLINE=99192826; PubMed=10092866;
RA   Bof M., Brandolin G., Satre M., Klein G.;
RT   "The mitochondrial adenine nucleotide translocator from Dictyostelium
RT   discoideum. Functional characterization and DNA sequencing.";
RL   Eur. J. Biochem. 259:795-800(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
DR   EMBL; AF100676; AAC79081.1; -.
DR   EMBL; AF039211; AAC77879.1; -.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR002067; Mit_carrier.
DR   Pfam; PF00153; mito_carr; 3.
DR   PRINTS; PR00926; MITOCARRIER.
DR   PROSITE; PS00215; MITOCH_CARRIER; 3.
KW   Inner membrane; Mitochondrion; Repeat; Transmembrane; Transport.
SQ   SEQUENCE   309 AA;  33468 MW;  D623DDBD4BA49474 CRC64;


  Query Match             57.4%;  Score 35;  DB 5;  Length 309;
  Best Local Similarity   70.0%;  Pred. No. 1e+02;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LATGSARQLT 11
           : |||||| |
Db     148 VGTGSARQFT 157


RESULT  12
Q9RWU3
ID   Q9RWU3      PRELIMINARY;      PRT;   423 AA.
AC   Q9RWU3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
```

```
DE   SPORULATION PROTEIN SPOIID-RELATED PROTEIN.
GN   DR0572.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001915; AAF10150.1; -.
DR   TIGR; DR0572; -.
KW   Complete proteome.
SQ   SEQUENCE   423 AA;  43100 MW;  D42323D17EF0E081 CRC64;


  Query Match             57.4%;  Score 35;  DB 16;  Length 423;
  Best Local Similarity   70.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
           |:| |:||||
Db     305 ASGRAQQLTF 314


RESULT  13
O95186
ID   O95186      PRELIMINARY;      PRT;   641 AA.
AC   O95186;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FATTY ACID TRANSPORT PROTEIN.
GN   FATP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE=99096471; PubMed=9878842;
RA   Fitscher B.A., Riedel H.D., Young K.C., Stremmel W.;
RT   "Tissue distribution and cDNA cloning of a human fatty acid transport
RT   protein (hsFATP4).";
RL   Biochim. Biophys. Acta 1443:381-385(1998).
DR   EMBL; AF055899; AAD11623.1; -.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00501; AMP-binding; 2.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ   SEQUENCE   641 AA;  71431 MW;  8C24F76C9BF81378 CRC64;


  Query Match             57.4%;  Score 35;  DB 4;  Length 641;
  Best Local Similarity   58.3%;  Pred. No. 2.1e+02;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     168 CLTTSRARALVF 179


RESULT  14
Q9C1M9
ID   Q9C1M9      PRELIMINARY;      PRT;   660 AA.
AC   Q9C1M9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA-L-RHAMNOSIDASE A PRECURSOR (EC 3.2.1.40).
GN   RHAA.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21218536; PubMed=11319105;
RA   Manzanares P., van Den Broeck H.C., de Graaff L.H., Visser J.;
RT   "Purification and Characterization of Two Different alpha-L-
RT   Rhamnosidases, RhaA and RhaB, from Aspergillus aculeatus.";
RL   Appl. Environ. Microbiol. 67:2230-2234(2001).
DR   EMBL; AF284761; AAK16249.1; -.
KW   Signal; Hydrolase; Glycosidase.
FT   SIGNAL        1      19       POTENTIAL.
SQ   SEQUENCE   660 AA;  71218 MW;  A550A9F4D0E42984 CRC64;


  Query Match             57.4%;  Score 35;  DB 3;  Length 660;
  Best Local Similarity   70.0%;  Pred. No. 2.1e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
           |||: :||||
Db     604 ATGTFQQLTF 613


RESULT  15
Q27966
ID   Q27966      PRELIMINARY;      PRT;  1028 AA.
AC   Q27966;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MYOSIN I.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ADRENAL GLAND;
RX   MEDLINE=94148088; PubMed=8313976;
RA   Zhu T., Ikebe M.;
RT   "Cloning of myosin I from bovine adrenal gland.";
RL   FEBS Lett. 339:31-36(1994).
DR   EMBL; U03420; AAA17565.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
SQ   SEQUENCE   1028 AA;  118020 MW;  5CEFEE28CF2A52D7 CRC64;


  Query Match             57.4%;  Score 35;  DB 6;  Length 1028;
  Best Local Similarity   58.3%;  Pred. No. 3.3e+02;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
```

Db 463 CLRPGEATDLTF 474

Search completed: June 6, 2002, 06:22:11
Job time: 380 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:20 ; Search time 52.04 Seconds
(without alignments)
25.613 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	100.0	12	19	AAW47591	T-cell receptor CD
2	50	82.0	267	19	AAW47588	T-cell receptor al
3	42.5	69.7	13	19	AAW47590	T-cell receptor CD
4	41	67.2	153	17	AAR98466	Alpha chain of sol
5	41	67.2	263	17	AAR97725	D10 single chain T
6	39	63.9	179	21	AAB25390	Pinus radiata cell
7	37.5	61.5	215	22	ABG11979	Novel human diagno
8	37	60.7	119	22	ABG23742	Novel human diagno
9	37	60.7	740	22	AAG65913	Amino acid sequenc
10	37	60.7	932	22	AAE05252	Mouse Nope (neighb
11	37	60.7	1250	22	AAG65914	Amino acid sequenc
12	37	60.7	1252	22	AAE05251	Mouse Nope (neighb
13	37	60.7	1395	22	ABG29676	Novel human diagno
14	37	60.7	1426	21	AAB15983	E. coli proliferat
15	36	59.0	108	21	AAB32964	Pinus radiata tran
16	36	59.0	217	22	AAU09176	Penicillium griseo
17	36	59.0	391	22	AAU71891	C. glutamicum meta
18	36	59.0	391	22	AAG91291	C glutamicum prote
19	36	59.0	391	22	AAB79663	Corynebacterium gl
20	36	59.0	391	22	AAB79783	Corynebacterium gl
21	36	59.0	391	22	AAB79797	Corynebacterium gl
22	36	59.0	406	22	ABG19026	Novel human diagno
23	36	59.0	428	18	AAW16543	50K-cellulase from
24	35	57.4	23	22	ABB31905	Peptide #4556 enco
25	35	57.4	23	22	ABB37143	Peptide #4649 enco
26	35	57.4	23	22	ABB22456	Protein #4455 enco
27	35	57.4	23	22	AAM70281	Human bone marrow
28	35	57.4	23	22	AAM18108	Peptide #4542 enco
29	35	57.4	23	22	AAM30620	Peptide #4657 enco
30	35	57.4	23	22	AAM05743	Peptide #4425 enco
31	35	57.4	511	21	AAY71058	Human membrane tra
32	35	57.4	511	22	AAB94242	Human protein sequ
33	35	57.4	616	21	AAB42756	Human ORFX ORF2520
34	35	57.4	632	22	AAB83236	Human FATP4 SEQ ID
35	35	57.4	632	22	AAB83240	Human FATP4 SEQ ID
36	35	57.4	643	20	AAY14943	Amino acid sequenc
37	35	57.4	643	20	AAY14949	Amino acid sequenc
38	35	57.4	643	22	AAB83233	Human FATP4 SEQ ID
39	35	57.4	643	22	AAB83242	Human FATP4 SEQ ID
40	35	57.4	643	22	AAB83249	Human FATP4 SEQ ID
41	35	57.4	935	22	ABG10172	Novel human diagno
42	35	57.4	1050	22	ABG10171	Novel human diagno
43	35	57.4	1063	22	AAM39991	Human polypeptide
44	34	55.7	61	22	AAU46008	Propionibacterium
45	34	55.7	170	21	AAY99741	Corn PI metabolism

ALIGNMENTS

RESULT 1
AAW47591
ID AAW47591 standard; peptide; 12 AA.
XX
AC AAW47591;
XX
DT 26-JUN-1998 (first entry)
XX
DE T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3.
XX
OS Homo sapiens.
XX
PN DE19625191-A1.
XX
PD 02-JAN-1998.
XX
PF 24-JUN-1996; 96DE-1025191.
XX
PR 24-JUN-1996; 96DE-1025191.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Schendel D;
XX
DR WPI; 1998-053442/06.
DR N-PSDB; AAV18708.
XX
PT Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX

```
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   12 AA;


  Query Match             100.0%;  Score 61;  DB 19;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           ||||||||||||
Db       1 clatgsarqltf 12


RESULT   2
AAW47588
ID   AAW47588 standard; Protein; 267 AA.
XX
AC   AAW47588;
XX
DT   26-JUN-1998  (first entry)
XX
DE   T-cell receptor alpha-chain.
XX
KW   Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma.
XX
OS   Homo sapiens.
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   N-PSDB; AAV18705.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Pages 11-13; 30pp; German.
XX
CC   The present sequence is the alpha-chain of a human T-cell
CC   receptor (TCR), which can be used in the diagnosis, monitoring,
CC   prevention and therapy of a tumour disease, specifically renal
CC   cell carcinoma.
XX
SQ   Sequence   267 AA;


  Query Match             82.0%;  Score 50;  DB 19;  Length 267;
  Best Local Similarity   83.3%;  Pred. No. 0.17;
  Matches   10;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           ||  ||||||||
Db     105 clvggsarqltf 116


RESULT   3
AAW47590
ID   AAW47590 standard; peptide; 13 AA.
XX
AC   AAW47590;
XX
DT   26-JUN-1998  (first entry)
XX
DE   T-cell receptor CDR3 alpha-region.
XX
KW   Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW   prevention; therapy; tumour disease; renal cell carcinoma;
KW   CDR3.
XX
OS   Homo sapiens.
XX
PN   DE19625191-A1.
XX
PD   02-JAN-1998.
XX
PF   24-JUN-1996;   96DE-1025191.
XX
PR   24-JUN-1996;   96DE-1025191.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Schendel D;
XX
DR   WPI; 1998-053442/06.
DR   N-PSDB; AAV18707.
XX
PT   Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT   diagnosis or therapy, especially of renal cell carcinoma
XX
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   13 AA;


  Query Match             69.7%;  Score 42.5;  DB 19;  Length 13;
  Best Local Similarity   76.9%;  Pred. No. 0.18;
  Matches   10;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 CLA-TGSARQLTF 12
           ||  :|||||||||
Db       1 clvlsgsarqltf 13


RESULT   4
AAR98466
ID   AAR98466 standard; Protein; 153 AA.
XX
AC   AAR98466;
XX
DT   13-JAN-1997  (first entry)
XX
DE   Alpha chain of soluble secreted D10 dual chain T-cell receptor.
XX
KW   T-cell receptor; TCR; pathogenic T-cell; killer T-cell;
KW   T-lymphocyte; inhibition; activation; graft rejection; antigen;
KW   diabetes; assay; detection; heterodimer; alpha chain; beta chain.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   Region          147
FT                   /label= C-alpha
XX
PN   WO9621028-A2.
XX
```

```
PD   11-JUL-1996.
XX
PF   28-DEC-1995;   95WO-US16937.
XX
PR   03-JAN-1995;   95US-0367589.
XX
PA   (PROC-) PROCEPT INC.
XX
PI   Banerji J,  Brauer P,  Jesson M,  Jones B,  Khandekar S;
PI   Mckeever U,  Naylor J;
XX
DR   WPI; 1996-334003/33.
DR   N-PSDB; AAT31598.
XX
PT   Soluble hetero:dimeric T cell receptor contg. linked alpha and beta
PT   sub:unit - useful to reduce pathogenic T cell activation, and to
PT   immunise mammals, e.g. to prevent graft rejection
XX
PS   Claim 1; Figure 1; 120pp; English.
XX
CC   A soluble heterodimeric T-cell receptor (TCR) which contains an
CC   alpha subunit connected by disulphide bonding to a beta subunit can
CC   be used to detect and analyse the peptide and MHC/HLA molecular
CC   constituents of TCR ligands. The TCR is useful to reduce the
CC   activation of pathogenic T cells in a mammal, and to immunise
CC   against TCR antigenic structures on the surface of such cells e.g to
CC   prevent graft rejection.  It can also be used to deplete pathogenic
CC   T-cells or inhibit their activation in a female's offspring by
CC   administration to the female during gestation or before weaning.
CC   The TCR and antibodies directed against it can also be used to
CC   deplete peripheral lymphocyte T-cells involved in T-cell mediated
CC   disease in a mammal or its offspring, particularly diabetes.  The
CC   heterodimeric TCR can also be used in an assay to detect pathogenic
CC   T-cells.
XX
SQ   Sequence   153 AA;


  Query Match             67.2%;  Score 41;  DB 17;  Length 153;
  Best Local Similarity   66.7%;  Pred. No. 4.8;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           | ||||  :|||
Db     118 caatgsfnkltf 129


RESULT   5
AAR97725
ID   AAR97725 standard; Protein; 263 AA.
XX
AC   AAR97725;
XX
DT   18-SEP-1996  (first entry)
XX
DE   D10 single chain T-cell receptor.
XX
KW   Single chain T-cell receptor; TCR; maltose binding protein;
KW   MBP-D10 scTCR; fusion protein; V-alpha; V-beta; antibody; vaccine;
KW   conalbumin.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   6..7
FT                   /note= "thrombin cleavage site"
FT   Domain          7..118
FT                   /label= V-beta
FT                   /note= "residue 8 (Ala in the wild-type) is
FT                          substituted by Ser to facilitate
FT                          thrombin cleavage"
FT   Region          119..145
FT                   /label= Linker
FT   Domain          146..257
FT                   /label= V-alpha
FT   Domain          258..263
FT                   /label= Hexahistidine_tail
XX
PN   WO9613593-A2.
XX
PD   09-MAY-1996.
XX
PF   26-OCT-1995;   95WO-US13770.
XX
PR   06-JUN-1995;   95US-0468131.
PR   26-OCT-1994;   94US-0329310.
PR   01-DEC-1994;   94US-0347893.
XX
PA   (PROC-) PROCEPT INC.
XX
PI   Banerji J,  Bettencourt B,  Dwyer D,  Jesson M,  Jones B;
PI   Khandekar S, McKeever U,  Naylor J;
XX
DR   WPI; 1996-239502/24.
DR   N-PSDB; AAT29757.
XX
PT   New fusion proteins comprising a single chain T-cell receptor  -
PT   used to develop prods. for use in detection, diagnosis, functional
PT   studies and therapy involving immune responses
XX
PS   Example 1; Page 61-62; 104pp; English.
XX
CC   A soluble single chain T-cell receptor (scTCR) (AAR97725) comprises
CC   the V-beta region of conalbumin-specific D10 T-cell receptor (ATCC
CC   TIB 224) joined to the D10 V-alpha region via a linker peptide.  It
CC   is obtd. by PCR amplification of V-beta and V-alpha sequences, and
CC   cloning into a vector that encodes the linker.  The entire scTCR
CC   sequence (AAT29757) is then cloned into vector pPR998 which encodes
CC   maltose binding protein (MBP).  The MBP-D10 scTCR fusion protein
CC   is expressed in E. coli and purified by affinity chromatography.
CC   Thrombin cleavage yields a soluble scTCR useful in detection,
CC   diagnosis, functional studies and therapy involving immune responses.
XX
SQ   Sequence   263 AA;


  Query Match             67.2%;  Score 41;  DB 17;  Length 263;
  Best Local Similarity   66.7%;  Pred. No. 8.6;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           | ||||  :|||
Db     235 caatgsfnkltf 246


RESULT   6
AAB25390
ID   AAB25390 standard; Protein; 179 AA.
XX
AC   AAB25390;
XX
DT   27-NOV-2000  (first entry)
XX
DE   Pinus radiata cell signalling involved protein SEQ ID NO:709.
XX
KW   Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW   plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW   environmental change; development; cell proliferation; differentiation;
KW   elongation; survival; disease resistance; nutrient metabolism.
XX
OS   Pinus radiata.
XX
PN   WO200042171-A1.
XX
PD   20-JUL-2000.
XX
```

PF 11-JAN-2000; 2000WO-US00724.
XX
PR 12-JAN-1999; 99US-0228986.
PR 01-NOV-1999; 99US-0162866.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Strabala TJ, Nieuwenhuizen NJ;
XX
DR WPI; 2000-476052/41.
XX
PT Isolated polynucleotide encoding a polypeptide involved in cell
PT signaling used for generating transgenic plants with modified responses
PT to external signals -
XX
PS Claim 3; Page 326; 527pp; English.
XX
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
XX
SQ Sequence 179 AA;

Query Match 63.9%; Score 39; DB 21; Length 179;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|:| |||| |::
Db 156 cialgsarglsy 167

RESULT 7
ABG11979
ID ABG11979 standard; Protein; 215 AA.
XX
AC ABG11979;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #11970.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS76166.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 42338; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215 AA;

Query Match 61.5%; Score 37.5; DB 22; Length 215;
Best Local Similarity 69.2%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 CLATG-SARQLTF 12
|||:| | |:|||
Db 81 clasgeslrkltf 93

RESULT 8
ABG23742
ID ABG23742 standard; Protein; 119 AA.
XX
AC ABG23742;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23733.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;

```
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87929.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 54101; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   119 AA;

  Query Match             60.7%;  Score 37;  DB 22;  Length 119;
  Best Local Similarity   63.6%;  Pred. No. 21;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLT 11
           || ||||  :|
Db      52 clmtgsaenvt 62


RESULT   9
AAG65913
ID   AAG65913 standard; protein; 740 AA.
XX
AC   AAG65913;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Amino acid sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdoch PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   N-PSDB; AAI67203.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer -
XX
PS   Claim 1; Page 89-91; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence   740 AA;

  Query Match             60.7%;  Score 37;  DB 22;  Length 740;
  Best Local Similarity   77.8%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |:|| ||||
Db     214 cvatnsarq 222


RESULT  10
AAE05252
ID   AAE05252 standard; Protein; 932 AA.
XX
AC   AAE05252;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) extracellular domain.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; extracellular domain.
XX
OS   Mus musculus.
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
```

```
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   N-PSDB; AAD10022.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 2; Page 81-83; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) extracellular domain.
XX
SQ   Sequence   932 AA;

  Query Match             60.7%;  Score 37;  DB 22;  Length 932;
  Best Local Similarity   77.8%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |:|| |||| 
Db     190 cvatnsarq 198


RESULT  11
AAG65914
ID   AAG65914 standard; protein; 1250 AA.
XX
AC   AAG65914;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Amino acid sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdoch PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   N-PSDB; AAI67204.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer  -
XX
PS   Claim 1; Page 91-94; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence   1250 AA;

  Query Match             60.7%;  Score 37;  DB 22;  Length 1250;
  Best Local Similarity   77.8%;  Pred. No. 2.6e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |:|| |||| 
Db     212 cvatnsarq 220


RESULT  12
AAE05251
ID   AAE05251 standard; Protein; 1252 AA.
XX
AC   AAE05251;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) protein.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Signal_peptide
FT   Protein         22..1252
FT                   /note= "Mature human Nope protein"
FT   Domain          22..953
FT                   /label= Extracellular_domain
FT   Domain          47..127
FT                   /label= Ig1
FT                   /note= "Immunoglobulin domain 1"
FT   Domain          155..218
FT                   /label= Ig2
```

```
FT                      /note= "Immunoglobulin domain 2"
FT   Domain             256..318
FT                      /label= Ig3
FT                      /note= "Immunoglobulin domain 3"
FT   Domain             347..411
FT                      /label= Ig4
FT                      /note= "Immunoglobulin domain 4"
FT   Domain             429..511
FT                      /label= Fn1
FT                      /note= "Fibronectin-type III domain 1"
FT   Domain             527..609
FT                      /label= Fn2
FT                      /note= "Fibronectin-type III domain 2"
FT   Domain             630..725
FT                      /label= Fn3
FT                      /note= "Fibronectin-type III domain 3"
FT   Domain             750..831
FT                      /label= Fn4
FT                      /note= "Fibronectin-type III domain 4"
FT   Domain             848..931
FT                      /label= Fn5
FT                      /note= "Fibronectin-type III domain 5"
FT   Domain             978..1252
FT                      /label= Intracellular_domain
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   N-PSDB; AAD10021.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 1; Fig 2B; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) protein.
XX
SQ   Sequence   1252 AA;
```

```
Query Match             60.7%;  Score 37;  DB 22;  Length 1252;
Best Local Similarity   77.8%;  Pred. No. 2.6e+02;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
             |:|| ||||
Db     211 cvatnsarq 219
```

```
RESULT  13
ABG29676
ID   ABG29676 standard; Protein; 1395 AA.
XX
AC   ABG29676;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29667.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS93863.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 60035; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1395 AA;
```

```
Query Match             60.7%;  Score 37;  DB 22;  Length 1395;
Best Local Similarity   70.0%;  Pred. No. 3e+02;
Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
```

```
          |||| ||:|:
Db    211 atgstrqmtw 220


RESULT 14
AAB15983
ID   AAB15983 standard; Protein; 1426 AA.
XX
AC   AAB15983;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated protein sequence SEQ ID NO:340.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   N-PSDB; AAA65988.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 11; Page 253-256; 316pp; English.
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
SQ   Sequence   1426 AA;

  Query Match             60.7%;  Score 37;  DB 21;  Length 1426;
  Best Local Similarity   70.0%;  Pred. No. 3e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLTF 12
           |||| ||:|:
Db     528 atgstrqmtw 537


RESULT 15
AAB32964
ID   AAB32964 standard; Protein; 108 AA.
XX
AC   AAB32964;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Pinus radiata transcription factor protein sequence #91.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB.
XX
OS   Pinus radiata.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 8; Page 371; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is one such
CC   transcription factor. The transcription factor may be used to produce a
CC   plant having modified gene expression such as a woody plant e.g. a
CC   eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
CC   to modify the activity of a polypeptide in a plant. The transcription
CC   factors of the present invention are members from the following families
CC   of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
CC   helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
CC   zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
CC   Cys2His2, CCAAT box elements and MYB.
XX
SQ   Sequence   108 AA;

  Query Match             59.0%;  Score 36;  DB 21;  Length 108;
  Best Local Similarity   63.6%;  Pred. No. 30;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LATGSARQLTF 12
           :|  ||||:||
Db      11 ianasarqvtf 21


Search completed: June  6, 2002, 06:17:21
Job time: 125 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:47 ; Search time 21.84 Seconds
(without alignments)
13.421 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	59.0	428	4	US-09-329-350-33	Sequence 33, Appl
2	35	57.4	632	4	US-09-232-200-34	Sequence 34, Appl
3	35	57.4	632	4	US-09-232-200-39	Sequence 39, Appl
4	35	57.4	632	4	US-09-232-197-34	Sequence 34, Appl
5	35	57.4	632	4	US-09-232-197-39	Sequence 39, Appl
6	35	57.4	632	4	US-09-232-201-34	Sequence 34, Appl
7	35	57.4	632	4	US-09-232-201-39	Sequence 39, Appl
8	35	57.4	643	4	US-09-232-200-27	Sequence 27, Appl
9	35	57.4	643	4	US-09-232-200-41	Sequence 41, Appl
10	35	57.4	643	4	US-09-232-200-53	Sequence 53, Appl
11	35	57.4	643	4	US-09-232-197-27	Sequence 27, Appl
12	35	57.4	643	4	US-09-232-197-41	Sequence 41, Appl
13	35	57.4	643	4	US-09-232-197-53	Sequence 53, Appl
14	35	57.4	643	4	US-09-232-201-27	Sequence 27, Appl
15	35	57.4	643	4	US-09-232-201-41	Sequence 41, Appl
16	35	57.4	643	4	US-09-232-201-53	Sequence 53, Appl
17	33	54.1	364	4	US-09-293-322C-7	Sequence 7, Appli
18	33	54.1	446	1	US-07-952-800-4	Sequence 4, Appli
19	33	54.1	456	2	US-08-709-979A-1	Sequence 1, Appli
20	33	54.1	456	3	US-08-709-974A-11	Sequence 11, Appl
21	33	54.1	496	4	US-09-147-009-8	Sequence 8, Appli
22	33	54.1	506	4	US-09-232-191-9	Sequence 9, Appli
23	33	54.1	506	4	US-09-232-200-9	Sequence 9, Appli
24	33	54.1	506	4	US-09-232-200-95	Sequence 95, Appl
25	33	54.1	506	4	US-09-232-197-9	Sequence 9, Appli
26	33	54.1	506	4	US-09-232-197-95	Sequence 95, Appl
27	33	54.1	506	4	US-09-232-201-9	Sequence 9, Appli
28	33	54.1	506	4	US-09-232-201-95	Sequence 95, Appl
29	33	54.1	525	4	US-08-764-870-7	Sequence 7, Appli
30	33	54.1	525	4	US-08-980-115-7	Sequence 7, Appli
31	33	54.1	533	1	US-07-952-800-2	Sequence 2, Appli
32	33	54.1	632	4	US-09-232-200-35	Sequence 35, Appl
33	33	54.1	632	4	US-09-232-197-35	Sequence 35, Appl
34	33	54.1	632	4	US-09-232-201-35	Sequence 35, Appl
35	33	54.1	643	4	US-09-232-200-42	Sequence 42, Appl
36	33	54.1	643	4	US-09-232-200-45	Sequence 45, Appl
37	33	54.1	643	4	US-09-232-200-71	Sequence 71, Appl
38	33	54.1	643	4	US-09-232-197-42	Sequence 42, Appl
39	33	54.1	643	4	US-09-232-197-45	Sequence 45, Appl
40	33	54.1	643	4	US-09-232-197-71	Sequence 71, Appl
41	33	54.1	643	4	US-09-232-201-42	Sequence 42, Appl
42	33	54.1	643	4	US-09-232-201-45	Sequence 45, Appl
43	33	54.1	643	4	US-09-232-201-71	Sequence 71, Appl
44	33	54.1	2813	3	US-08-896-449A-2	Sequence 2, Appli
45	33	54.1	2813	3	US-09-132-652-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-329-350-33
; Sequence 33, Application US/09329350
; Patent No. 6184019
; GENERAL INFORMATION:
;   APPLICANT: Miettinen-Oinonen, Arja
;   APPLICANT: Londesborough, John
;   APPLICANT: Vehmaanper , Jari
;   APPLICANT: Haakana, Heli
;   APPLICANT: M ntyl , Arja
;   APPLICANT: Lantto, Raija
;   APPLICANT: Elovainio, Minna
;   APPLICANT: Joutsjoki, Vesa
;   APPLICANT: Paloheimo, Marja
;   APPLICANT: Suominen, Pirkko
;   TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
;   TITLE OF INVENTION: USES THEREOF
;   NUMBER OF SEQUENCES: 45
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
;     STREET: 1100 New York Avenue, N.W., Suite 600
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.50 inch
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/329,350
;     FILING DATE: Herewith
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/841,636
;     FILING DATE: 30-APR-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/005,335
;     FILING DATE: 17-OCT-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/007,926
;     FILING DATE: 04-DEC-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/020,840
;     FILING DATE: 28-JUN-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/732,181
;     FILING DATE: 16-OCT-1996
;   PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  PCT/FI96/00550
;      FILING DATE:  17-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shea Jr., Timothy
;      REGISTRATION NUMBER:  41,306
;      REFERENCE/DOCKET NUMBER:  1716.0510006/MAC/TJS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  428 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Melanocarpus albomyces
;      STRAIN:  ALKO4237
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..428
;      OTHER INFORMATION:  /label= 50K-cellulase
US-09-329-350-33


  Query Match           59.0%;  Score 36;  DB 4;  Length 428;
  Best Local Similarity  70.0%;  Pred. No. 25;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQL 10
           | |||:|| |
Db     332 CAATGAARYL 341


RESULT   2
US-09-232-200-34
; Sequence 34, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-200-34


  Query Match           57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity  58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     157 CLTTSRARALVF 168


RESULT   3
US-09-232-200-39
; Sequence 39, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-200-39


  Query Match           57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity  58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     157 CLTTSRARALVF 168


RESULT   4
US-09-232-197-34
; Sequence 34, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-197-34


  Query Match           57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity  58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
```

```
Db    157 CLTTSRARALVF 168


RESULT   5
US-09-232-197-39
; Sequence 39, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-197-39


  Query Match             57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity   58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     157 CLTTSRARALVF 168


RESULT   6
US-09-232-201-34
; Sequence 34, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MC
;  CURRENT APPLICATION NUMBER: US/09/232,201A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-201-34


  Query Match             57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity   58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     157 CLTTSRARALVF 168


RESULT   7
US-09-232-201-39
; Sequence 39, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MC
;  CURRENT APPLICATION NUMBER: US/09/232,201A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-201-39


  Query Match             57.4%;  Score 35;  DB 4;  Length 632;
  Best Local Similarity   58.3%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     157 CLTTSRARALVF 168


RESULT   8
US-09-232-200-27
; Sequence 27, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27
;   LENGTH: 643
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-200-27
```

```
  Query Match             57.4%;  Score 35;  DB 4;  Length 643;
  Best Local Similarity   58.3%;  Pred. No. 61;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     168 CLTTSRARALVF 179


RESULT   9
US-09-232-200-41
; Sequence 41, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
;   LENGTH: 643
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-200-41


  Query Match             57.4%;  Score 35;  DB 4;  Length 643;
  Best Local Similarity   58.3%;  Pred. No. 61;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     168 CLTTSRARALVF 179


RESULT  10
US-09-232-200-53
; Sequence 53, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
;   LENGTH: 643
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-200-53


  Query Match             57.4%;  Score 35;  DB 4;  Length 643;
  Best Local Similarity   58.3%;  Pred. No. 61;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     168 CLTTSRARALVF 179


RESULT  11
US-09-232-197-27
; Sequence 27, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27
;   LENGTH: 643
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-232-197-27


  Query Match             57.4%;  Score 35;  DB 4;  Length 643;
  Best Local Similarity   58.3%;  Pred. No. 61;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           || |  || | |
Db     168 CLTTSRARALVF 179


RESULT  12
US-09-232-197-41
; Sequence 41, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
```

; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-41

Query Match 57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|| | || | |
Db 168 CLTTSRARALVF 179

RESULT 13
US-09-232-197-53
; Sequence 53, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-53

Query Match 57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|| | || | |
Db 168 CLTTSRARALVF 179

RESULT 14
US-09-232-201-27
; Sequence 27, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-201-27

Query Match 57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|| | || | |
Db 168 CLTTSRARALVF 179

RESULT 15
US-09-232-201-41
; Sequence 41, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-201-41

Query Match 57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
|| | || | |
Db 168 CLTTSRARALVF 179

Search completed: June 6, 2002, 06:15:48
Job time: 32 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:25:52 ; Search time 14.16 Seconds
(without alignments)
81.432 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1579

Minimum DB seq length: 0
Maximum DB seq length: 12

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	37.7	10	2	S23371	T-cell receptor al
2	22	36.1	12	2	PH1611	Ig H chain V-D-J r
3	21	34.4	8	2	PT0639	T-cell receptor be
4	21	34.4	10	2	PQ0783	NADH dehydrogenase
5	21	34.4	10	2	S65728	hemoglobin, extrac
6	21	34.4	10	2	PT0212	T-cell receptor al
7	20	32.8	6	2	PT0568	T-cell receptor be
8	20	32.8	8	2	S08996	hypertrehalosemic
9	20	32.8	8	2	B49823	adipokinetic hormo
10	20	32.8	8	2	B44960	neuropeptide Led-C
11	20	32.8	8	2	A33995	adipokinetic hormo
12	20	32.8	8	2	S11545	adipokinetic hormo
13	20	32.8	9	2	A24244	adipokinetic hormo
14	20	32.8	9	2	A29477	diuretic neuropept
15	20	32.8	10	2	JC1416	hypertrehalosemic
16	20	32.8	10	2	S09138	hypertrehalosemic
17	20	32.8	10	2	A31571	hypertrehalosemic/
18	20	32.8	10	2	B33995	hypotrehalosemic h
19	19	31.1	8	2	I48934	apolipoprotein A-I
20	19	31.1	10	2	PH0895	T-cell receptor be
21	19	31.1	12	2	PH1185	T-cell receptor al
22	19	31.1	12	2	A37933	Ig lambda chain J
23	19	31.1	12	2	PH1454	T-cell receptor al
24	18	29.5	9	2	PH0918	T-cell receptor be
25	18	29.5	10	2	H60588	sperm-activating p
26	18	29.5	10	2	S70251	nitrogenase (EC 1.
27	18	29.5	11	2	PH1583	Ig H chain V-D-J r
28	18	29.5	11	2	PT0214	T-cell receptor be
29	18	29.5	12	2	S36902	Em protein - wheat
30	17	27.9	5	2	B44817	34.5K structural p
31	17	27.9	5	2	D44817	35K structural pro
32	17	27.9	7	2	S45311	microcin C7 - Esch
33	17	27.9	7	2	B34818	vicilin 57K chain
34	17	27.9	8	2	PH0934	T-cell receptor be
35	17	27.9	9	2	PH0943	T-cell receptor be
36	17	27.9	10	2	S48182	bacterioferritin -
37	17	27.9	10	2	PH0946	T-cell receptor be
38	17	27.9	11	2	PH0940	T-cell receptor be
39	17	27.9	12	2	A49033	T-cell receptor de
40	17	27.9	12	2	A53252	pollen major aller
41	16	26.2	6	2	S14159	parasporal crystal
42	16	26.2	7	2	PN0150	omega-gliadine 1'
43	16	26.2	9	2	JQ0914	MHC class I histoc
44	16	26.2	9	2	A28924	fructose-bisphosph
45	16	26.2	10	2	E61512	variant surface gl

ALIGNMENTS

RESULT 1
S23371
T-cell receptor alpha chain J region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
C;Accession: S23371
R;Pluschke, G.; Ricken, G.; Taube, H.; Kroninger, S.; Melchers, I.; Peter, H.H.; Eich
Eur. J. Immunol. 21, 2749-2754, 1991
A;Title: Biased T cell receptor V(alpha) region repertoire in the synovial fluid of r
A;Reference number: S23364; MUID:92037820
A;Accession: S23371
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-10 <PLU>
A;Cross-references: EMBL:X58166
C;Keywords: T-cell receptor

Query Match 37.7%; Score 23; DB 2; Length 10;
Best Local Similarity 50.0%; Pred. No. 4e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy       3 ATGSARQLTF 12
           |:| | : ||
Db       1 ASGEAGKSTF 10
```

RESULT 2
PH1611
Ig H chain V-D-J region (wild-type clone 341) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 17-Mar-1999
C;Accession: PH1611
R;Levinson, D.A.; Campos-Torres, J.; Leder, P.
J. Exp. Med. 178, 317-329, 1993
A;Title: Molecular characterization of transgene-induced immunodeficiency in B-less m
A;Reference number: PH1580; MUID:93301609
A;Accession: PH1611
A;Molecule type: DNA
A;Residues: 1-12 <LEV>
A;Experimental source: bone marrow pre-B lymphocyte
C;Keywords: immunoglobulin

Query Match 36.1%; Score 22; DB 2; Length 12;
Best Local Similarity 33.3%; Pred. No. 7.6e+02;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```
Qy       1 CLATGSARQLTF 12
           |   || | : :
Db       1 CARQGSYRAMDY 12
```

```
RESULT   3
PT0639
T-cell receptor beta chain V-D-J region (111-1AA) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0639
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0639
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-8 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor

  Query Match             34.4%;  Score 21;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 2.8e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSAR 8
           :|| ||
Db       3 STGGAR 8


RESULT   4
PQ0783
NADH dehydrogenase (EC 1.6.99.3) 30K chain - fava bean mitochondrion (fragment)
N;Alternate names: complex I 30K chain; NADH--ubiquinone reductase 30K chain
C;Species: mitochondrion Vicia faba (fava bean)
C;Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C;Accession: PQ0783
R;Leterme, S.; Boutry, M.
Plant Physiol. 102, 435-443, 1993
A;Title: Purification and preliminary characterization of mitochondrial complex I (NADH:
A;Reference number: PQ0775; MUID:94151437
A;Accession: PQ0783
A;Molecule type: protein
A;Residues: 1-10 <LET>
C;Comment: Complex I, mitochondrial NADH-ubiquiquinone reductase, is the first of the th
 ranging from 5K to 75K.
C;Comment: This enzyme catalyzes electron transfer from endogenous NADH to ubiquinone by
C;Genetics:
A;Genome: mitochondrion
C;Keywords: electron transfer; mitochondrion; oxidoreductase

  Query Match             34.4%;  Score 21;  DB 2;  Length 10;
  Best Local Similarity   44.4%;  Pred. No. 1e+03;
  Matches    4;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 ATGSARQLT 11
           || :|: :|
Db       1 ATEAAKHIT 9


RESULT   5
S65728
hemoglobin, extracellular, chain d1 - earthworm (Lumbricus terrestris) (fragment)
C;Species: Lumbricus terrestris (common earthworm)
C;Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
C;Accession: S65728
R;Fushitani, K.; Higashiyama, K.; Asao, M.; Hosokawa, K.
Biochim. Biophys. Acta 1292, 273-280, 1996
A;Title: Characterization of the constituent polypeptides of the extracellular hemoglobi
A;Reference number: S65721; MUID:96176855
A;Accession: S65728
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-10 <FUS>

  Query Match             34.4%;  Score 21;  DB 2;  Length 10;
  Best Local Similarity   50.0%;  Pred. No. 1e+03;
  Matches    4;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSAR 8
           || | | :
Db       2 CLVTESLK 9


RESULT   6
PT0212
T-cell receptor alpha chain V-J region (4-1-E.2) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 30-May-1997
C;Accession: PT0212
R;Nakano, N.; Kikutani, H.; Nishimoto, H.; Kishimoto, T.
J. Exp. Med. 173, 1091-1097, 1991
A;Title: T cell receptor V gene usage of islet beta cell-reactive T cells is not rest
A;Reference number: PT0209; MUID:91217621
A;Accession: PT0212
A;Molecule type: mRNA
A;Residues: 1-10 <NAK>
C;Keywords: T-cell receptor

  Query Match             34.4%;  Score 21;  DB 2;  Length 10;
  Best Local Similarity   40.0%;  Pred. No. 1e+03;
  Matches    4;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQL 10
           |   | | :|
Db       1 CAVAGGADRL 10


RESULT   7
PT0568
T-cell receptor beta chain V-D-J region (141-1C) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0568
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0568
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor

  Query Match             32.8%;  Score 20;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.8e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATGSAR 8
           |:| ||
Db       1 ASGDAR 6


RESULT   8
S08996
hypertrehalosemic hormone II - oriental cockroach
N;Alternate names: Pea-CAH-II
C;Species: Blatta orientalis (oriental cockroach)
C;Date: 30-Jun-1992 #sequence_revision 24-Oct-1997 #text_change 31-Oct-1997
C;Accession: S08996
R;Gaede, G.; Rinehart, K.L.
Biol. Chem. Hoppe-Seyler 371, 345-354, 1990
A;Title: Primary structures of hypertrehalosaemic neuropeptides isolated from the cor
```

entalis and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombard
A;Reference number: S08995; MUID:90253659
A;Accession: S08996
A;Molecule type: protein
A;Residues: 1-8 <GAE>
A;Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 9
B49823
adipokinetic hormone II - American cockroach
N;Alternate names: neuropeptide M-II; periplanetin CC-1
C;Species: Periplaneta americana (American cockroach)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: B49823; A05170
R;Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller, C.
Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984
A;Title: Isolation and primary structure of two peptides with cardioacceleratory and hyp
A;Reference number: A49823; MUID:84298179
A;Accession: B49823
A;Molecule type: protein
A;Residues: 1-8 <SCA>
R;Witten, J.L.; Schaffer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr., K.L
Biochem. Biophys. Res. Commun. 124, 350-358, 1984
A;Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment mas
A;Reference number: A90118; MUID:85046530
A;Accession: A05170
A;Molecule type: protein
A;Residues: 'E',2-8 <WIT>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 10
B44960
neuropeptide Led-CC-II - Colorado potato beetle
C;Species: Leptinotarsa decemlineata (Colorado potato beetle)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: B44960
R;Gaede, G.; Kellner, R.
Peptides 10, 1287-1289, 1989
A;Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and
A;Reference number: A44960; MUID:90160053
A;Accession: B44960
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimenta

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 11
A33995
adipokinetic hormone - black horse fly
C;Species: Tabanus atratus (black horse fly)
C;Date: 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change 31-Oct-1997
C;Accession: A33995
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhan
Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989
A;Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotre
A;Reference number: A33995; MUID:90046758
A;Accession: A33995
A;Molecule type: protein
A;Residues: 1-8 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted
F;8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 12
S11545
adipokinetic hormone - nestling-sucking blowfly
C;Species: Protophormia terraenovae (nestling-sucking blowfly)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 31-Oct-1997
C;Accession: S11545
R;Gaede, G.; Wilps, H.; Kellner, R.
Biochem. J. 269, 309-313, 1990
A;Title: Isolation and structure of a novel charged member of the red-pigment-concent
erraenovae (Diptera).
A;Reference number: S11545; MUID:90351345
A;Accession: S11545
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 13
A24244
adipokinetic hormone - bollworm
N;Alternate names: Hez-AKH

C;Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)
C;Date: 31-Mar-1988 #sequence_revision 23-Mar-1995 #text_change 31-Oct-1997
C;Accession: A24244
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridgway
Biochem. Biophys. Res. Commun. 135, 622-628, 1986
A;Title: Isolation and primary structure of a peptide from the corpora cardiaca of Helic
A;Reference number: A24244; MUID:86186794
A;Accession: A24244
A;Molecule type: protein
A;Residues: 1-9 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 14
A29477
diuretic neuropeptide F1 - migratory locust
C;Species: Locusta migratoria (migratory locust)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 31-Dec-1993
C;Accession: A29477
R;Proux, J.P.; Miller, C.A.; Li, J.P.; Carney, R.L.; Girardie, A.; Delaage, M.; Schooley
Biochem. Biophys. Res. Commun. 149, 180-186, 1987
A;Title: Identification of an arginine vasopressin-like diuretic hormone from Locusta mi
A;Reference number: A29477; MUID:88077077
A;Accession: A29477
A;Molecule type: protein
A;Residues: 1-9 <PRO>
A;Note: two neuropeptides, F1 and F2, were identified. F2 is an antiparallel dimer of F1
C;Keywords: neuropeptide

Query Match 32.8%; Score 20; DB 2; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CLATGSAR 8
|| | |
Db 1 CLITNCPR 8

RESULT 15
JC1416
hypertrehalosemic hormone I - stick insect (Carausius morosus)
N;Alternate names: neuropeptide Cam-HrTH-I
N;Contains: hypertrehalosemic factor II
C;Species: Carausius morosus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: JC1416; S07157
R;Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.
Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992
A;Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick in
A;Reference number: JC1416; MUID:93129188
A;Accession: JC1416
A;Molecule type: protein
A;Residues: 1-10 <GAE1>
R;Gaede, G.; Rinehart Jr., K.L.
Biol. Chem. Hoppe-Seyler 368, 67-75, 1987
A;Title: Primary structure of the hypertrehalosaemic factor II from the corpus cardiacum
A;Reference number: S07157; MUID:87157103
A;Accession: S07157
A;Molecule type: protein
A;Residues: 'Z',2-10 <GAE2>
C;Comment: Hypertrehalosemic factor II lacks the tryptophan modification.
C;Comment: This peptide raises hemolymph levels of trehalose in the cockroach Peripla
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Binding site: carbohydrate (Trp) (covalent) #status experimental
F;10/Modified site: amidated carboxyl end (Thr) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

Search completed: June 6, 2002, 06:27:51
Job time: 119 sec

entalis and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombard
A;Reference number: S08995; MUID:90253659
A;Accession: S08996
A;Molecule type: protein
A;Residues: 1-8 <GAE>
A;Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 9
B49823
adipokinetic hormone II - American cockroach
N;Alternate names: neuropeptide M-II; periplanetin CC-1
C;Species: Periplaneta americana (American cockroach)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: B49823; A05170
R;Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller, C.
Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984
A;Title: Isolation and primary structure of two peptides with cardioacceleratory and hyp
A;Reference number: A49823; MUID:84298179
A;Accession: B49823
A;Molecule type: protein
A;Residues: 1-8 <SCA>
R;Witten, J.L.; Schaffer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr., K.L
Biochem. Biophys. Res. Commun. 124, 350-358, 1984
A;Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment mas
A;Reference number: A90118; MUID:85046530
A;Accession: A05170
A;Molecule type: protein
A;Residues: 'E',2-8 <WIT>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 10
B44960
neuropeptide Led-CC-II - Colorado potato beetle
C;Species: Leptinotarsa decemlineata (Colorado potato beetle)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: B44960
R;Gaede, G.; Kellner, R.
Peptides 10, 1287-1289, 1989
A;Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and
A;Reference number: A44960; MUID:90160053
A;Accession: B44960
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimenta

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 11
A33995
adipokinetic hormone - black horse fly
C;Species: Tabanus atratus (black horse fly)
C;Date: 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change 31-Oct-1997
C;Accession: A33995
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhan
Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989
A;Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotre
A;Reference number: A33995; MUID:90046758
A;Accession: A33995
A;Molecule type: protein
A;Residues: 1-8 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted
F;8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 12
S11545
adipokinetic hormone - nestling-sucking blowfly
C;Species: Protophormia terraenovae (nestling-sucking blowfly)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 31-Oct-1997
C;Accession: S11545
R;Gaede, G.; Wilps, H.; Kellner, R.
Biochem. J. 269, 309-313, 1990
A;Title: Isolation and structure of a novel charged member of the red-pigment-concent
erraenovae (Diptera).
A;Reference number: S11545; MUID:90351345
A;Accession: S11545
A;Molecule type: protein
A;Residues: 1-8 <GAE>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 13
A24244
adipokinetic hormone - bollworm
N;Alternate names: Hez-AKH

C;Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)
C;Date: 31-Mar-1988 #sequence_revision 23-Mar-1995 #text_change 31-Oct-1997
C;Accession: A24244
R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridgway
Biochem. Biophys. Res. Commun. 135, 622-628, 1986
A;Title: Isolation and primary structure of a peptide from the corpora cardiaca of Helic
A;Reference number: A24244; MUID:86186794
A;Accession: A24244
A;Molecule type: protein
A;Residues: 1-9 <JAF>
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 14
A29477
diuretic neuropeptide F1 - migratory locust
C;Species: Locusta migratoria (migratory locust)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 31-Dec-1993
C;Accession: A29477
R;Proux, J.P.; Miller, C.A.; Li, J.P.; Carney, R.L.; Girardie, A.; Delaage, M.; Schooley
Biochem. Biophys. Res. Commun. 149, 180-186, 1987
A;Title: Identification of an arginine vasopressin-like diuretic hormone from Locusta mi
A;Reference number: A29477; MUID:88077077
A;Accession: A29477
A;Molecule type: protein
A;Residues: 1-9 <PRO>
A;Note: two neuropeptides, F1 and F2, were identified. F2 is an antiparallel dimer of F1
C;Keywords: neuropeptide

Query Match 32.8%; Score 20; DB 2; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CLATGSAR 8
|| | |
Db 1 CLITNCPR 8

RESULT 15
JC1416
hypertrehalosemic hormone I - stick insect (Carausius morosus)
N;Alternate names: neuropeptide Cam-HrTH-I
N;Contains: hypertrehalosemic factor II
C;Species: Carausius morosus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: JC1416; S07157
R;Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.
Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992
A;Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick in
A;Reference number: JC1416; MUID:93129188
A;Accession: JC1416
A;Molecule type: protein
A;Residues: 1-10 <GAE1>
R;Gaede, G.; Rinehart Jr., K.L.
Biol. Chem. Hoppe-Seyler 368, 67-75, 1987
A;Title: Primary structure of the hypertrehalosaemic factor II from the corpus cardiacum
A;Reference number: S07157; MUID:87157103
A;Accession: S07157
A;Molecule type: protein
A;Residues: 'Z',2-10 <GAE2>
C;Comment: Hypertrehalosemic factor II lacks the tryptophan modification.
C;Comment: This peptide raises hemolymph levels of trehalose in the cockroach Peripla
C;Superfamily: adipokinetic hormone
C;Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide; pyroglutamic
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;8/Binding site: carbohydrate (Trp) (covalent) #status experimental
F;10/Modified site: amidated carboxyl end (Thr) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

Search completed: June 6, 2002, 06:27:51
Job time: 119 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:26:37 ; Search time 10.17 Seconds
(without alignments)
45.687 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 463

Minimum DB seq length: 0
Maximum DB seq length: 12

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	20	32.8	8	1	AKH_TABAT	P14595	tabanus atr
2	20	32.8	8	1	HTF2_PERAM	P04549	periplaneta
3	20	32.8	9	1	DNF1_LOCMI	P16339	locusta mig
4	20	32.8	10	1	HTF2_CARMO	P11385	carausius m
5	20	32.8	10	1	HTF_HELZE	P16353	heliothis z
6	20	32.8	10	1	HTF_TABAT	P14596	tabanus atr
7	17	27.9	10	1	AKHX_LOCMI	P81626	locusta mig
8	16	26.2	12	1	CALM_TETTH	Q05055	tetrahymena
9	15	24.6	7	1	UH11_RAT	P56576	rattus norv
10	15	24.6	8	1	HTF_TENMO	P25419	tenebrio mo
11	15	24.6	8	1	RPCH_PANBO	P08939	pandalus bo
12	15	24.6	8	1	UPAA_HUMAN	P30096	homo sapien
13	15	24.6	9	1	AL11_CARMA	P81814	carcinus ma
14	15	24.6	9	1	LITR_PHYRO	P08946	phyllomedus
15	15	24.6	10	1	GON2_CHEPR	P80678	chelyosoma
16	15	24.6	10	1	VEG6_BACSU	P80699	bacillus su
17	15	24.6	11	1	CA41_LITCI	P82091	litoria cit
18	15	24.6	11	1	CA42_LITCI	P82092	litoria cit
19	15	24.6	11	1	RANC_RANPI	P08951	rana pipien
20	14	23.0	8	1	ALL4_CYDPO	P82155	cydia pomon
21	14	23.0	9	1	HUTU_KLEAE	P12381	klebsiella
22	14	23.0	9	1	XYLA_STRSQ	P19149	streptomyce
23	14	23.0	10	1	GAJU_HUMAN	P01358	homo sapien
24	14	23.0	10	1	URE3_MORMO	P17339	morganella
25	14	23.0	11	1	MORN_HUMAN	P01163	homo sapien
26	13	21.3	8	1	LPMS_STAEP	P23211	staphylococ
27	13	21.3	9	1	FAR9_ASCSU	P43172	ascaris suu
28	13	21.3	9	1	LMT3_LOCMI	P41489	locusta mig
29	13	21.3	10	1	MOSQ_CLYJA	P19962	clypeaster
30	13	21.3	10	1	TKU1_UREUN	P40751	urechis uni
31	13	21.3	12	1	TIN2_HOPTI	P82652	hoplobatrac
32	12	19.7	8	1	AKHG_GRYBI	P14086	gryllus bim
33	12	19.7	8	1	AKH_LIBAU	P25418	libellula a
34	12	19.7	8	1	AKH_MELML	P25423	melolontha
35	12	19.7	8	1	HTF1_PERAM	P04548	periplaneta
36	12	19.7	9	1	CONO_CONST	P05487	conus stria
37	12	19.7	9	1	DSIP_RABIT	P01158	oryctolagus
38	12	19.7	10	1	FARC_CALVO	P41867	calliphora
39	12	19.7	10	1	HTF1_ROMMI	P18110	romalea mic
40	12	19.7	10	1	HTF_NAUCI	P10939	nauphoeta c
41	12	19.7	10	1	PSBF_CAPAN	Q03367	capsicum an
42	12	19.7	10	1	RCA_PINPS	P81084	pinus pinas
43	12	19.7	11	1	CA21_LITCI	P82087	litoria cit
44	12	19.7	11	1	CA22_LITCI	P82088	litoria cit
45	12	19.7	11	1	TIN4_HOPTI	P82654	hoplobatrac

ALIGNMENTS

RESULT 1
AKH_TABAT
ID AKH_TABAT STANDARD; PRT; 8 AA.
AC P14595;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Adipokinetic hormone (AKH) (Dipteran corpora cardiaca factor I)
DE (DCC I).
OS Tabanus atratus (Horse fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;
OC Tabanidae; Tabanus.
OX NCBI_TaxID=7207;
RN [1]
RP SEQUENCE.
RC TISSUE=Corpora cardiaca;
RX MEDLINE=90046758; PubMed=2813385;
RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,
RA Vogel V.W., Zhang Y.-S., Hayes D.K.;
RT "Primary structure of two neuropeptide hormones with adipokinetic and
RT hypotrehalosemic activity isolated from the corpora cardiaca of horse
RT flies (Diptera).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).
CC -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR PIR; A33995; A33995.
DR InterPro; IPR002047; AKH.
DR PROSITE; PS00256; AKH; 1.
KW Neuropeptide; Amidation; Flight.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 8 8 AMIDATION.
SQ SEQUENCE 8 AA; 949 MW; 86786771A9D1A736 CRC64;

Query Match 32.8%; Score 20; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLTF 12
||||
Db 1 QLTF 4

RESULT 2
HTF2_PERAM
ID HTF2_PERAM STANDARD; PRT; 8 AA.
AC P04549;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hypertrehalosaemic factor II (Neuropeptide M-II) (Periplanetin CC-2)
DE (PeA-CAH-II) (LeD-CC-II) (Hypertrehalosaemic neuropeptide II).

```
OS   Periplaneta americana (American cockroach),
OS   Leptinotarsa decemlineata (Colorado potato beetle), and
OS   Blatta orientalis (Oriental cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
OX   NCBI_TaxID=6978, 7539, 6976;
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.americana;
RX   MEDLINE=85046530; PubMed=6548628;
RA   Witten J.L., Schaffer M.H., O'Shea M., Cook J.C., Hemling M.E.,
RA   Rinehart K.L. Jr.;
RT   "Structures of two cockroach neuropeptides assigned by fast atom
RT   bombardment mass spectrometry.";
RL   Biochem. Biophys. Res. Commun. 124:350-358(1984).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.americana;
RX   MEDLINE=84298179; PubMed=6591205;
RA   Scarborough R.M., Jamieson G.C., Kalish F., Kramer S.J., McEnroe G.A.,
RA   Miller C.A., Schooley D.A.;
RT   "Isolation and primary structure of two peptides with
RT   cardioacceleratory and hyperglycemic activity from the corpora
RT   cardiaca of Periplaneta americana.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:5575-5579(1984).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.decemlineata; TISSUE=Corpora cardiaca;
RX   MEDLINE=90160053; PubMed=2576128;
RA   Gaede G., Kellner R.;
RT   "The metabolic neuropeptides of the corpus cardiacum from the potato
RT   beetle and the American cockroach are identical.";
RL   Peptides 10:1287-1289(1989).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=B.orientalis; TISSUE=Corpora cardiaca;
RX   MEDLINE=90253659; PubMed=2340112;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structures of hypertrehalosaemic neuropeptides isolated from
RT   the corpora cardiaca of the cockroaches Leucophaea maderae,
RT   Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT   and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT   atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 371:345-354(1990).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A05170; A05170.
DR   PIR; S08996; S08996.
DR   PIR; B44960; B44960.
DR   PIR; B49823; B49823.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1006 MW;  86745771A9D1A736 CRC64;
```

```
  Query Match             32.8%;  Score 20;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           ||||
Db       1 QLTF 4
```

```
RESULT   3
DNF1_LOCMI
ID   DNF1_LOCMI      STANDARD;       PRT;      9 AA.
AC   P16339;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Locupressin (Diuretic neuropeptide F1/F2).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Suboesophageal ganglion, and Thoracic ganglion;
RX   MEDLINE=88077077; PubMed=3689410;
RA   Proux J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,
RA   Delaage M., Schooley D.A.;
RT   "Identification of an arginine vasopressin-like diuretic hormone from
RT   Locusta migratoria.";
RL   Biochem. Biophys. Res. Commun. 149:180-186(1987).
CC   -!- FUNCTION: DIURETIC HORMONE.
CC   -!- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF F1.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A29477; A29477.
DR   InterPro; IPR000981; Neurhypophys_horm.
DR   Pfam; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Neuropeptide; Amidation.
FT   DISULFID      1      6       IN F1.
FT   DISULFID      1      1       INTERCHAIN (WITH C-6') (IN F2).
FT   DISULFID      6      6       INTERCHAIN (WITH C-1') (IN F2).
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  976 MW;  56EB176EB451A057 CRC64;
```

```
  Query Match             32.8%;  Score 20;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1e+05;
  Matches    4;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CLATGSAR 8
           || |   |
Db       1 CLITNCPR 8
```

```
RESULT   4
HTF2_CARMO
ID   HTF2_CARMO      STANDARD;       PRT;     10 AA.
AC   P11385;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hypertrehalosaemic factor II (HTF-II) (HRTH-II) (Hypertrehalosaemic
DE   neuropeptide II).
OS   Carausius morosus (Indian stick insect), and
OS   Extatosoma tiaratum (Stick insect).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Phasmatodea; Heteronemiidae;
OC   Carausius.
OX   NCBI_TaxID=7022, 7024;
RN   [1]
RP   SEQUENCE.
RC   SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX   MEDLINE=87157103; PubMed=3828078;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structure of the hypertrehalosaemic factor II from the
RT   corpus cardiacum of the Indian stick insect, Carausius morosus,
RT   determined by fast atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 368:67-75(1987).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=E.tiaratum; TISSUE=Corpora cardiaca;
RX   MEDLINE=90253659; PubMed=2340112;
RA   Gaede G., Rinehart K.L. Jr.;
RT   "Primary structures of hypertrehalosaemic neuropeptides isolated from
```

```
RT   the corpora cardiaca of the cockroaches Leucophaea maderae,
RT   Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT   and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT   atom bombardment mass spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 371:345-354(1990).
RN   [3]
RP   CARBOHYDRATE-LINKAGE SITE.
RC   SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX   MEDLINE=93129188; PubMed=1482345;
RA   Gaede G., Kellner R., Rinehart K.L. Jr., Proefke M.L.;
RT   "A tryptophan-substituted member of the AKH/RPCH family isolated from
RT   a stick insect corpus cardiacum.";
RL   Biochem. Biophys. Res. Commun. 189:1303-1309(1992).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- MASS SPECTROMETRY: MW=1308.61; METHOD=FAB.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S07157; S07157.
DR   PIR; S09138; S09138.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Glycoprotein.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   CARBOHYD      8      8       C-LINKED (MAN) (PROBABLE).
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1164 MW;  9B9036745771A9D1 CRC64;
```

```
  Query Match             32.8%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           ||||
Db       1 QLTF 4
```

```
RESULT   5
HTF_HELZE
ID   HTF_HELZE      STANDARD;      PRT;     10 AA.
AC   P16353;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic hormone (HeZ-HRTH).
OS   Heliothis zea (Corn earworm) (Bollworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
OX   NCBI_TaxID=7113;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=88326324; PubMed=3415690;
RA   Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Bird T.G.,
RA   Tseng C.M., Zhang Y.S., Hayes D.K.;
RT   "Isolation and primary structure of a neuropeptide hormone from
RT   Heliothis zea with hypertrehalosemic and adipokinetic activities.";
RL   Biochem. Biophys. Res. Commun. 155:344-350(1988).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A31571; A31571.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1096 MW;  8E70367865A5B9D1 CRC64;
```

```
  Query Match             32.8%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           ||||
Db       1 QLTF 4
```

```
RESULT   6
HTF_TABAT
ID   HTF_TABAT      STANDARD;      PRT;     10 AA.
AC   P14596;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic factor (HOTH) (Dipteran corpora cardiaca factor II)
DE   (DCC II).
OS   Tabanus atratus (Horse fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;
OC   Tabanidae; Tabanus.
OX   NCBI_TaxID=7207;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=90046758; PubMed=2813385;
RA   Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,
RA   Vogel V.W., Zhang Y.-S., Hayes D.K.;
RT   "Primary structure of two neuropeptide hormones with adipokinetic and
RT   hypotrehalosemic activity isolated from the corpora cardiaca of horse
RT   flies (Diptera).";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; B33995; B33995.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1169 MW;  916036786771A9D1 CRC64;
```

```
  Query Match             32.8%;  Score 20;  DB 1;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           ||||
Db       1 QLTF 4
```

```
RESULT   7
AKHX_LOCMI
ID   AKHX_LOCMI     STANDARD;      PRT;     10 AA.
AC   P81626;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Peptide hormone.
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RA   Siegert K.J.;
RL   Submitted (DEC-1998) to the SWISS-PROT data bank.
```

```
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE REGULATION OF LOCUST
CC       INTERMEDIARY METABOLISM, BEHAVIOR AND/OR DEVELOPMENT.
CC   -!- SIMILARITY: SOME SIMILARITY TO THE AKH / HRTH / RPCH FAMILY.
DR   InterPro; IPR002047; AKH.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      10     10       AMIDATION.
SQ   SEQUENCE   10 AA;  1222 MW;  81BFF67AB415B9D1 CRC64;


  Query Match             27.9%;  Score 17;  DB 1;  Length 10;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           |:||
Db       1 QVTF 4


RESULT   8
CALM_TETTH
ID   CALM_TETTH     STANDARD;      PRT;    12 AA.
AC   Q05055;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Calmodulin (Fragment).
OS   Tetrahymena thermophila.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281388; PubMed=8506136;
RA   Katoh M., Hirono M., Takemasa T., Kimura M., Watanabe Y.;
RT   "A micronucleus-specific sequence exists in the 5'-upstream region of
RT   calmodulin gene in Tetrahymena thermophila.";
RL   Nucleic Acids Res. 21:2409-2414(1993).
CC   -!- FUNCTION: CALMODULIN MEDIATES THE CONTROL OF A LARGE NUMBER OF
CC       ENZYMES BY CA(++). AMONG THE ENZYMES TO BE STIMULATED BY THE
CC       CALMODULIN-CA(++) COMPLEX ARE A NUMBER OF PROTEIN KINASES AND
CC       PHOSPHATASES.
CC   -!- MISCELLANEOUS: THIS PROTEIN HAS FOUR FUNCTIONAL CALCIUM-BINDING
CC       SITES.
CC   -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12774; BAA02239.1; -.
DR   HSSP; P02593; 2CLN.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; PARTIAL.
KW   Calcium-binding; Repeat; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1393 MW;  83F31CD443DB1B01 CRC64;


  Query Match             26.2%;  Score 16;  DB 1;  Length 12;
  Best Local Similarity   80.0%;  Pred. No. 3.3e+03;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 ARQLT 11
           | |||
Db       1 ADQLT 5


RESULT   9
UH11_RAT
ID   UH11_RAT       STANDARD;      PRT;     7 AA.
AC   P56576;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Unknown protein from 2D-page of heart tissue (Spot P11) (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=WISTAR; TISSUE=Heart;
RA   Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA   Jungblut P.R.;
RL   Submitted (SEP-1998) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 8.5, ITS MW IS: 42 kDa.
FT   UNSURE        2      2       OR A.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  775 MW;  6866DB040DC5A6B0 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 SARQ 9
           |||:
Db       2 SARE 5


RESULT  10
HTF_TENMO
ID   HTF_TENMO      STANDARD;      PRT;     8 AA.
AC   P25419;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Hypertrehalosaemic factor (HOTH) (Hypertrehalosemic neuropeptide).
OS   Tenebrio molitor (Yellow mealworm), and
OS   Zophobas rugipes.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Tenebrionidae; Tenebrio.
OX   NCBI_TaxID=7067, 7075;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Corpora cardiaca;
RX   MEDLINE=90341081; PubMed=2381871;
RA   Gaede G., Rosinski G.;
RT   "The primary structure of the hypertrehalosemic neuropeptide from
RT   tenebrionid beetles: a novel member of the AKH/RPCH family.";
RL   Peptides 11:455-459(1990).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A43976; A43976.
DR   PIR; B43976; B43976.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1005 MW;  86745775B9C44736 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1e+05;
```

```
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           || |
Db       1 QLNF 4


RESULT  11
RPCH_PANBO
ID   RPCH_PANBO      STANDARD;      PRT;      8 AA.
AC   P08939;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Red pigment concentrating hormone (RPCH).
OS   Pandalus borealis (Northern red shrimp).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea; Pandaloidea;
OC   Pandalidae; Pandalus.
OX   NCBI_TaxID=6703;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=75054965; PubMed=4433569;
RA   Fernlund P.;
RT   "Structure of the red-pigment-concentrating hormone of the shrimp,
RT   Pandalus borealis.";
RL   Biochim. Biophys. Acta 371:304-311(1974).
CC   -!- FUNCTION: THIS HORMONE ADAPTS THE ANIMAL TO LIGHT BACKGROUNDS BY
CC       STIMULATING CONCENTRATION OF THE PIGMENT OF ITS RED BODY-
CC       CHROMATOPHORES.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S07139; S07139.
DR   InterPro; IPR002047; AKH.
DR   PROSITE; PS00256; AKH; 1.
KW   Pigment; Hormone; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  948 MW;  86786775B9C44736 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 QLTF 12
           || |
Db       1 QLNF 4


RESULT  12
UPAA_HUMAN
ID   UPAA_HUMAN      STANDARD;      PRT;      8 AA.
AC   P30096;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Unknown protein from 2D-page of plasma (Spot 36) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Plasma;
RX   MEDLINE=93092937; PubMed=1459097;
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 7, ITS MW IS: 12 kDa.
DR   SWISS-2DPAGE; P30096; HUMAN.
FT   NON_TER       1      1
FT   VARIANT       5      5       F -> P.
FT                                /FTId=VAR_000004.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  909 MW;  86677B59D1A72042 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 LTF 12
           |||
Db       3 LTF 5


RESULT  13
AL11_CARMA
ID   AL11_CARMA      STANDARD;      PRT;      9 AA.
AC   P81814;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Carcinustatin 11.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
OX   NCBI_TaxID=6759;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX   MEDLINE=98121193; PubMed=9461295;
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  927 MW;  832D79CDCB46D861 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 ATG 5
           |||
Db       1 ATG 3


RESULT  14
LITR_PHYRO
ID   LITR_PHYRO      STANDARD;      PRT;      9 AA.
AC   P08946;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Rhodei-litorin.
OS   Phyllomedusa rohdei (Rohde's leaf frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Phyllomedusa.
OX   NCBI_TaxID=8394;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85127560; PubMed=3838283;
RA   Barra D., Erspamer G.F., Simmaco M., Bossa F., Melchiorri P.,
RA   Erspamer V.;
RT   "Rohdei-litorin: a new peptide from the skin of Phyllomedusa rohdei.";
```

```
RL   FEBS Lett. 182:53-56(1985).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN
CC       FAMILY.
DR   PIR; S07241; S07241.
DR   InterPro; IPR000874; Bombesin.
DR   Pfam; PF02044; Bombesin; 1.
DR   PROSITE; PS00257; BOMBESIN; 1.
KW   Bombesin family; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1090 MW;  4ECCC1E861ADC377 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 ATG 5
          |||
Db      4 ATG 6


RESULT  15
GON2_CHEPR
ID   GON2_CHEPR     STANDARD;      PRT;    10 AA.
AC   P80678;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Gonadoliberin II (Gonadotropin-releasing hormone II) (GnRH-II)
DE   (Luliberin II).
OS   Chelyosoma productum.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC   Corellidae; Chelyosoma.
OX   NCBI_TaxID=71177;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=96413669; PubMed=8816823;
RA   Powell J.F.F., Reska-Skinner S.M., Prakash M.O., Fischer W.H.,
RA   Park M., Rivier J.E., Craig A.G., Mackie G.O., Sherwood N.M.;
RT   "Two new forms of gonadotropin-releasing hormone in a protochordate
RT   and the evolutionary implications.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:10461-10464(1996).
CC   -!- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND
CC       FOLLICLE-STIMULATING HORMONES.
CC   -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: GNRH NEURONS LIE WITHIN BLOOD SINUSES CLOSE TO
CC       THE GONODUCTS AND GONADS IN BOTH JUVENILES AND ADULTS, IMPLYING
CC       THAT THE NEUROPEPTIDE IS RELEASED INTO THE BLOODSTREAM.
CC   -!- MASS SPECTROMETRY: MW=1117.52; METHOD=MALDI.
CC   -!- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR   InterPro; IPR002012; GnRH.
DR   PROSITE; PS00473; GNRH; 1.
KW   Hormone; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID      6      6       INTERCHAIN.
FT   MOD_RES      10     10       AMIDATION (BY SIMILARITY).
SQ   SEQUENCE   10 AA;  1135 MW;  284B38D1EEB735A3 CRC64;


  Query Match             24.6%;  Score 15;  DB 1;  Length 10;
  Best Local Similarity   60.0%;  Pred. No. 4.3e+03;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CLATG 5
          | | |
Db      6 CHAPG 10


Search completed: June  6, 2002, 06:29:54
```

Job time: 197 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:27:32 ; Search time 23.79 Seconds
(without alignments)
87.261 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 1842

Minimum DB seq length: 0
Maximum DB seq length: 12

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	37.7	9	11	Q9QZA8	Q9qza8 mus musculu
2	21	34.4	8	11	P70243	P70243 mus musculu
3	21	34.4	12	10	Q945C3	Q945c3 crypthecodi
4	19	31.1	8	11	Q60615	Q60615 mus musculu
5	19	31.1	11	11	Q9R0K9	Q9r0k9 mus musculu
6	18	29.5	8	2	O09258	O09258 synechococc
7	18	29.5	10	11	Q9R0Z0	Q9r0z0 mus musculu
8	18	29.5	12	10	Q02319	Q02319 pinus sylve
9	18	29.5	12	10	Q02320	Q02320 pinus sylve
10	18	29.5	12	10	Q38715	Q38715 arachis hyp
11	17	27.9	7	2	Q47505	Q47505 escherichia
12	17	27.9	11	4	Q9UNL8	Q9unl8 homo sapien
13	17	27.9	12	12	O10421	O10421 influenza a
14	16	26.2	9	7	Q31415	Q31415 gallus gall
15	16	26.2	11	5	P82700	P82700 leucophaea
16	16	26.2	12	4	Q9UBJ5	Q9ubj5 homo sapien
17	16	26.2	12	10	Q9SYT4	Q9syt4 arabidopsis
18	16	26.2	12	10	Q41856	Q41856 zea mays (m
19	15	24.6	8	6	P82929	P82929 bos taurus
20	15	24.6	8	12	O89965	O89965 polyomaviru
21	15	24.6	9	12	Q9PYK1	Q9pyk1 simian viru
22	15	24.6	9	12	Q9IBM8	Q9ibm8 simian viru
23	15	24.6	10	10	P81899	P81899 prunus dulc
24	15	24.6	11	3	Q9UR95	Q9ur95 pichia angu
25	15	24.6	11	11	Q99JC3	Q99jc3 rattus sp.
26	15	24.6	11	12	O40974	O40974 cauliflower
27	15	24.6	11	13	Q90735	Q90735 gallus gall
28	15	24.6	12	2	P83054	P83054 bacteroides
29	15	24.6	12	3	Q9UR28	Q9ur28 filobasidie
30	15	24.6	12	3	Q9UR22	Q9ur22 cryptococcu
31	15	24.6	12	6	Q9TRU9	Q9tru9 bos taurus
32	15	24.6	12	9	Q9T1J4	Q9t1j4 bacteriopha
33	15	24.6	12	10	P82329	P82329 pisum sativ
34	15	24.6	12	12	Q69232	Q69232 bovine herp
35	14	23.0	7	13	O42564	O42564 fugu rubrip
36	14	23.0	8	2	Q44463	Q44463 rhizobiacea
37	14	23.0	9	6	Q9XT05	Q9xt05 macropus ru
38	14	23.0	10	5	P82222	P82222 bombyx mori
39	14	23.0	11	5	P82698	P82698 leucophaea
40	14	23.0	11	6	Q9TRX0	Q9trx0 sus scrofa
41	14	23.0	11	7	O77913	O77913 oreochromis
42	14	23.0	11	11	Q9QXM6	Q9qxm6 mus musculu
43	14	23.0	11	11	Q60807	Q60807 mus musculu
44	14	23.0	12	2	Q9R7F1	Q9r7f1 staphylococ
45	14	23.0	12	6	Q9XT42	Q9xt42 canis famil

ALIGNMENTS

RESULT 1
Q9QZA8
ID Q9QZA8 PRELIMINARY; PRT; 9 AA.
AC Q9QZA8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C-TYPE LECTIN DCL1 (FRAGMENT).
GN DCL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Gorski K., Huang X., Tseng S.-Y., Rattis F., Pardoll D., Tsuchiya H.;
RT "Dendritic cell regulation of DCL1 mRNA expression.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF192526; AAF04843.1; -.
DR MGD; MGI:2136650; Dcl1.
KW Lectin.
FT NON_TER 9 9
SQ SEQUENCE 9 AA; 994 MW; 342161AB172EBAB7 CRC64;

Query Match 37.7%; Score 23; DB 11; Length 9;
Best Local Similarity 80.0%; Pred. No. 5.6e+05;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLATG 5
|| ||
Db 4 CLETG 8

RESULT 2
P70243
ID P70243 PRELIMINARY; PRT; 8 AA.
AC P70243;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)

```
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SKELETAL MUSCLE-SPECIFIC CALCIUM CHANNEL (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RA   Ophoff R.A.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X98325; CAA66969.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  865 MW;  D9C37DDB1861ADDE CRC64;


  Query Match             34.4%;  Score 21;  DB 11;  Length 8;
  Best Local Similarity   71.4%;  Pred. No. 5.6e+05;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 ATGSARQ 9
           ||| | |
Db       2 ATGEAWQ 8


RESULT   3
Q945C3
ID   Q945C3      PRELIMINARY;      PRT;    12 AA.
AC   Q945C3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   P80 PROTEIN (FRAGMENT).
OS   Crypthecodinium cohnii (Dinoflagellate).
OC   Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Crypthecodiniaceae;
OC   Crypthecodinium.
OX   NCBI_TaxID=2866;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99433238; PubMed=10505419;
RA   Ausseil J., Soyer-Gobillard M.O., Geraud M.L., Bhaud Y., Baines I.,
RA   Preston T., Moreau H.;
RT   "Characterization of p80, a novel nuclear and cytoplasmic protein in
RT   dinoflagellates.";
RL   Protist 150:197-211(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Guillebault D., Derelle E., Lozano J.C., Bingham S., Moreau H.;
RT   "A single TBP-like protein is present in the marine unicellular
RT   organism: the dinoflagellate Crypthecodinium cohnii.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF417568; AAL15906.1; -.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1364 MW;  615BF873FE204414 CRC64;


  Query Match             34.4%;  Score 21;  DB 10;  Length 12;
  Best Local Similarity   57.1%;  Pred. No. 2e+03;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 SARQLTF 12
           |||:: |
Db       3 SARRILF 9


RESULT   4
Q60615
ID   Q60615      PRELIMINARY;      PRT;     8 AA.
AC   Q60615;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   APOLIPOPROTEIN A-II (FRAGMENT).
GN   APOA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=94319082; PubMed=8043949;
RA   Ko M.S., Wang X., Horton J.H., Hagen M.D., Takahashi N., Maezaki Y.,
RA   Nadeau J.H.;
RT   "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";
RL   Mamm. Genome 5:349-355(1994).
DR   EMBL; U05691; AAB60462.1; -.
DR   MGD; MGI:88050; Apoa2.
KW   Lipoprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  720 MW;  783DDDC5B861AB18 CRC64;


  Query Match             31.1%;  Score 19;  DB 11;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 TGSA 7
           ||||
Db       3 TGSA 6


RESULT   5
Q9R0K9
ID   Q9R0K9      PRELIMINARY;      PRT;    11 AA.
AC   Q9R0K9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BCL-2 ASSOCIATED X PROTEIN (FRAGMENT).
GN   BAX OR BAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20035745; PubMed=10570968;
RA   Igata E., Inoue T., Ohtani-Fujita N., Sowa Y., Tsujimoto Y., Sakai T.;
RT   "Molecular cloning and functional analysis of the murine bax gene
RT   promoter.";
RL   Gene 238:407-415(1999).
DR   EMBL; AB029557; BAA82406.1; -.
DR   MGD; MGI:99702; Bax.
FT   NON_TER      11     11
SQ   SEQUENCE   11 AA;  1037 MW;  90AAB34E36DB1865 CRC64;


  Query Match             31.1%;  Score 19;  DB 11;  Length 11;
  Best Local Similarity   66.7%;  Pred. No. 4.5e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 GSARQL 10
           ||  ||
Db       3 GSGEQL 8


RESULT   6
O09258
ID   O09258      PRELIMINARY;      PRT;     8 AA.
AC   O09258;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE   NIFH (FRAGMENT).
GN   NIFH.
OS   Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothece PCC 8801).
OC   Bacteria; Cyanobacteria; Chroococcales; Cyanothece.
OX   NCBI_TaxID=41431;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RF-1;
RX   MEDLINE=99231861; PubMed=10217509;
RA   Huang T.C., Lin R.F., Chu M.K., Chen H.M.;
RT   "Organization and expression of nitrogen-fixation genes in the aerobic
RT   nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain
RT   RF-1.";
RL   Microbiology 145:743-753(1999).
DR   EMBL; AF001780; AAC33369.1; -.
FT   NON_TER       8       8
SQ   SEQUENCE   8 AA;  985 MW;  F16B59CDD046C406 CRC64;

  Query Match             29.5%;  Score 18;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 5.6e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 RQLTF 12
           ||: |
Db       2 RQIAF 6


RESULT   7
Q9R0Z0
ID   Q9R0Z0      PRELIMINARY;      PRT;    10 AA.
AC   Q9R0Z0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CW17 PROTEIN (FRAGMENT).
GN   CW17.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=LIVER;
RX   MEDLINE=99287587; PubMed=10360842;
RA   Wrehlke C., Wiedemeyer W.R., Schmitt-Wrede H.P., Mincheva A.,
RA   Lichter P., Wunderlich F.;
RT   "Genomic organization of mouse gene zfp162 (mzfm).";
RL   DNA Cell Biol. 18:419-428(1999).
DR   EMBL; Y14702; CAB45189.1; -.
FT   NON_TER      10      10
SQ   SEQUENCE   10 AA;  946 MW;  836D48ADD44DD861 CRC64;

  Query Match             29.5%;  Score 18;  DB 11;  Length 10;
  Best Local Similarity   60.0%;  Pred. No. 6.3e+03;
  Matches    3;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LATGS 6
           :|||:
Db       1 MATGA 5


RESULT   8
Q02319
ID   Q02319      PRELIMINARY;      PRT;    12 AA.
AC   Q02319;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CHALCONE SYNTHASE (FRAGMENT).
OS   Pinus sylvestris (Scots pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEEDLING;
RX   MEDLINE=93015948; PubMed=1400374;
RA   Schroder G., Schroder J.;
RT   "A single change of histidine to glutamine alters the substrate
RT   preference of a stilbene synthase.";
RL   J. Biol. Chem. 267:20558-20560(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEEDLING;
RX   MEDLINE=92163014; PubMed=1536925;
RA   Fliegmann J., Schroder G., Schanz S., Britsch L., Schroder J.;
RT   "Molecular analysis of chalcone and dihydropinosylvin synthase from
RT   Scots pine (Pinus sylvestris), and differential regulation of these
RT   and related enzyme activities in stressed plants.";
RL   Plant Mol. Biol. 18:489-503(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEEDLING;
RA   Schroeder J.;
RL   Submitted (AUG-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L00657; AAA50522.1; -.
FT   NON_TER      12      12
SQ   SEQUENCE   12 AA;  1303 MW;  DF8E5A1BE0CEB866 CRC64;

  Query Match             29.5%;  Score 18;  DB 10;  Length 12;
  Best Local Similarity   50.0%;  Pred. No. 7.6e+03;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGS 6
           | | |:
Db       7 CFAGGT 12


RESULT   9
Q02320
ID   Q02320      PRELIMINARY;      PRT;    12 AA.
AC   Q02320;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CHALCONE SYNTHASE (FRAGMENT).
OS   Pinus sylvestris (Scots pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93015948; PubMed=1400374;
RA   Schroder G., Schroder J.;
RT   "A single change of histidine to glutamine alters the substrate
RT   preference of a stilbene synthase.";
RL   J. Biol. Chem. 267:20558-20560(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92163014; PubMed=1536925;
RA   Fliegmann J., Schroder G., Schanz S., Britsch L., Schroder J.;
RT   "Molecular analysis of chalcone and dihydropinosylvin synthase from
RT   Scots pine (Pinus sylvestris), and differential regulation of these
RT   and related enzyme activities in stressed plants.";
RL   Plant Mol. Biol. 18:489-503(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Schroeder J.;
RL   Submitted (AUG-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L00658; AAA50523.1; -.
FT   NON_TER      12      12
SQ   SEQUENCE   12 AA;  1303 MW;  EDB35A1BE0CEB871 CRC64;
```

```
  Query Match             29.5%;  Score 18;  DB 10;  Length 12;
  Best Local Similarity   50.0%;  Pred. No. 7.6e+03;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGS 6
           | | |:
Db       7 CFAGGT 12


RESULT  10
Q38715
ID   Q38715      PRELIMINARY;      PRT;    12 AA.
AC   Q38715;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RESVERATROL SYNTHASE (FRAGMENT).
OS   Arachis hypogaea (Peanut).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC   Arachis.
OX   NCBI_TaxID=3818;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lanz T., Schroder G., Schroder J.;
RT   "Differential regulation of genes for resveratrol synthase in cell
RT   cultures of Arachis hypogaea L.";
RL   Planta 181:169-175(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91236783; PubMed=2033084;
RA   Schroeder J.;
RT   "The role of cysteines in polyketide synthases. Site-directed
RT   mutagenesis of resveratrol and chalcone synthases, two key enzymes in
RT   different plant specific pathways.";
RL   J. Biol. Chem. 266:9971-9976(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93015948; PubMed=1400374;
RA   Schroeder G., Schroeder J.;
RT   "A single change of histidine to glutamine alters the substrate
RT   preference of a stilbene synthase.";
RL   J. Biol. Chem. 267:20558-20560(1992).
DR   EMBL; L00953; AAC41685.1; -.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1294 MW;  C0B35A1BE0CEB866 CRC64;

  Query Match             29.5%;  Score 18;  DB 10;  Length 12;
  Best Local Similarity   50.0%;  Pred. No. 7.6e+03;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGS 6
           | | |:
Db       7 CFAGGT 12


RESULT  11
Q47505
ID   Q47505      PRELIMINARY;      PRT;     7 AA.
AC   Q47505;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MCCA PROTEIN.
GN   MCCA.
OS   Escherichia coli.
OG   Plasmid pMccC7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96099297; PubMed=8522520;
RA   Gonzalez-Pastor J.E., San Millan J.L., Castilla M.A., Moreno F.;
RT   "Structure and organization of plasmid genes required to produce the
RT   translation inhibitor microcin C7.";
RL   J. Bacteriol. 177:7131-7140(1995).
DR   EMBL; X57583; CAA40808.1; -.
KW   Plasmid.
SQ   SEQUENCE   7 AA;  763 MW;  644DD44861B406F0 CRC64;

  Query Match             27.9%;  Score 17;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 5.6e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LATGSA 7
           : ||:|
Db       1 MRTGNA 6


RESULT  12
Q9UNL8
ID   Q9UNL8      PRELIMINARY;      PRT;    11 AA.
AC   Q9UNL8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   APC2 PROTEIN (FRAGMENT).
GN   APC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Carr I.M., Markham A.F., Colleta P.L., Wai L., Askham J., Morrison E.,
RA   Meredith D.M.;
RT   "APC2 partial gene sequence.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF110338; AAD29275.1; -.
FT   NON_TER       1      1
FT   NON_TER      11     11
SQ   SEQUENCE   11 AA;  1326 MW;  75881D7BB441EAB4 CRC64;

  Query Match             27.9%;  Score 17;  DB 4;  Length 11;
  Best Local Similarity   75.0%;  Pred. No. 1.1e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLAT 4
           || |
Db       8 CLQT 11


RESULT  13
O10421
ID   O10421      PRELIMINARY;      PRT;    12 AA.
AC   O10421;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEURAMINIDASE (FRAGMENT).
OS   influenza A virus (A/South Carolina/1/18 (H1N1)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX   NCBI_TaxID=59375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/SOUTH CAROLINA/1/18 (H1N1);
RX   MEDLINE=97218304; PubMed=9065404;
RA   Taubenberger J.K., Reid A.H., Krafft A.E., Bijwaard K.E.,
RA   Fanning T.G.;
```

```
RT   "Initial genetic characterization of the 1918 'Spanish' influenza
RT   virus.";
RL   Science 275:1793-1796(1997).
DR   EMBL; U94893; AAC57065.1; -.
FT   NON_TER       1      1
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1248 MW;  D13CE0DB8F5862D2 CRC64;


  Query Match             27.9%;  Score 17;  DB 12;  Length 12;
  Best Local Similarity   40.0%;  Pred. No. 1.2e+04;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATG 5
           |:  |
Db       3 CMVVG 7


RESULT  14
Q31415
ID   Q31415      PRELIMINARY;      PRT;     9 AA.
AC   Q31415;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanki T., Kuwasawa N., Sekiya Y., Ichikawa Y.;
RT   "Responsive expression of a MHC class I epitope and genes following
RT   Marek's disease virus infection.";
RL   Submitted (JAN-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D90399; BAA14395.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  859 MW;  8A55A76455B861B5 CRC64;


  Query Match             26.2%;  Score 16;  DB 7;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 5.6e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 ATGS 6
           :|||
Db       2 STGS 5


RESULT  15
P82700
ID   P82700      PRELIMINARY;      PRT;    11 AA.
AC   P82700;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PERIVISCEROKININ-3 (LEM-PVK-3).
OS   Leucophaea maderae (Madeira cockroach),
OS   Nauphoeta cinerea (Cinereous cockroach) (Gray cockroach),
OS   Blaberus craniifer,
OS   Blaptica dubia (Argentinian wood cockroach), and
OS   Gromphadorina portentosa (Cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
OX   NCBI_TaxID=6988, 6990, 6982, 132935, 36953;
RN   [1]
RP   SEQUENCE, FUNCTION, AND MASS SPECTROSCOPY.
RC   TISSUE=ABDOMINAL PERISYMPATHETIC ORGANS;
RX   MEDLINE=20307624; PubMed=10849006;
RA   Predel R., Kellner R., Baggerman G., Steinmetzer T., Schoofs L.;
RT   "Identification of novel perivisceroki nins from single neurohaemal
RT   release sites in insects. MS/MS fragmentation complemented by Edman
RT   degradation.";
RL   Eur. J. Biochem. 267:3869-3873(2000).
CC   -!- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
CC       (MYOTROPIC ACTIVITY).
CC   -!- MASS SPECTROMETRY: MW=1146.6; METHOD=MALDI.
KW   Neuropeptide; Amidation.
FT   MOD_RES      11     11       AMIDATION.
SQ   SEQUENCE   11 AA;  1147 MW;  2F4D9FF2D7605698 CRC64;


  Query Match             26.2%;  Score 16;  DB 5;  Length 11;
  Best Local Similarity   37.5%;  Pred. No. 1.7e+04;
  Matches    3;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 GSARQLTF 12
           ||:  : |
Db       1 GSSGMIPF 8


Search completed: June  6, 2002, 06:30:44
Job time: 192 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:47 ; Search time 29.23 Seconds
(without alignments)
45.600 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 158732

Minimum DB seq length: 0
Maximum DB seq length: 12

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	100.0	12	19	AAW47591	T-cell receptor CD
2	28	45.9	9	15	AAR55332	Inhibitor of tyros
3	28	45.9	10	22	AAG95029	Human complementar
4	28	45.9	10	22	AAG95031	Human complementar
5	28	45.9	10	22	AAG95037	Human complementar
6	28	45.9	10	22	AAG95041	Human complementar
7	28	45.9	10	22	AAG95045	Human complementar
8	28	45.9	10	22	AAG95091	Human complementar
9	28	45.9	10	22	AAG95093	Human complementar
10	28	45.9	10	22	AAG95095	Human complementar
11	28	45.9	10	22	AAG95097	Human complementar
12	28	45.9	10	22	AAG95099	Human complementar
13	28	45.9	10	22	AAG95101	Human complementar
14	28	45.9	10	22	AAG95115	Human complementar
15	28	45.9	10	22	AAG95141	Human complementar
16	28	45.9	10	22	AAG95143	Human complementar
17	28	45.9	10	22	AAG95187	Human complementar
18	28	45.9	10	22	AAG95223	Human complementar
19	28	45.9	10	22	AAG95233	Human complementar
20	28	45.9	10	22	AAG95235	Human complementar
21	28	45.9	10	22	AAG95237	Human complementar
22	28	45.9	10	22	AAG95239	Human complementar
23	28	45.9	10	22	AAG95243	Human complementar
24	28	45.9	10	22	AAG95251	Human complementar
25	28	45.9	10	22	AAG95257	Human complementar
26	28	45.9	10	22	AAG95295	Human complementar
27	28	45.9	10	22	AAG95311	Human complementar
28	28	45.9	10	22	AAG95327	Human complementar
29	28	45.9	10	22	AAG95329	Human complementar
30	28	45.9	10	22	AAG95335	Human complementar
31	28	45.9	10	22	AAG95339	Human complementar
32	28	45.9	10	22	AAG95341	Human complementar
33	28	45.9	10	22	AAG95343	Human complementar
34	28	45.9	10	22	AAG95349	Human complementar
35	28	45.9	10	22	AAG95353	Human complementar
36	28	45.9	10	22	AAG95359	Human complementar
37	28	45.9	10	22	AAG95365	Human complementar
38	28	45.9	10	22	AAG95375	Human complementar
39	28	45.9	10	22	AAG95379	Human complementar
40	28	45.9	10	22	AAG95389	Human complementar
41	28	45.9	10	22	AAG95397	Human complementar
42	28	45.9	11	19	AAW64705	Seq ID 43 from WO9
43	27	44.3	10	22	AAG95027	Human complementar
44	27	44.3	10	22	AAG95087	Human complementar
45	27	44.3	10	22	AAG95089	Human complementar

ALIGNMENTS

RESULT 1
AAW47591
ID AAW47591 standard; peptide; 12 AA.
XX
AC AAW47591;
XX
DT 26-JUN-1998 (first entry)
XX
DE T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3.
XX
OS Homo sapiens.
XX
PN DE19625191-A1.
XX
PD 02-JAN-1998.
XX
PF 24-JUN-1996; 96DE-1025191.
XX
PR 24-JUN-1996; 96DE-1025191.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Schendel D;
XX
DR WPI; 1998-053442/06.
DR N-PSDB; AAV18708.
XX
PT Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX

```
PS   Example 1; Page 17; 30pp; German.
XX
CC   The present sequence is the CDR3 alpha-region of a human
CC   T-cell receptor (TCR), which can be used in the diagnosis,
CC   monitoring, prevention and therapy of a tumour disease,
CC   specifically renal cell carcinoma.
XX
SQ   Sequence   12 AA;


  Query Match             100.0%;  Score 61;  DB 19;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQLTF 12
           ||||||||||||
Db       1 clatgsarqltf 12


RESULT   2
AAR55332
ID   AAR55332 standard; peptide; 9 AA.
XX
AC   AAR55332;
XX
DT   30-JAN-1995  (first entry)
XX
DE   Inhibitor of tyrosine kinase.
XX
KW   PTK;  signal transduction; proliferation; cancer; psoriasis; viral;
KW   inflammatory; allergic; cardiovascular; diseases.
XX
OS   Synthetic.
XX
PN   WO9411392-A.
XX
PD   26-MAY-1994.
XX
PF   03-NOV-1993;   93WO-US10610.
XX
PR   06-NOV-1992;   92US-0973136.
PR   27-OCT-1993;   93US-0139913.
XX
PA   (WARN ) WARNER LAMBERT CO.
XX
PI   Dobrusin EM,  Fry DW,  Mcnamara DJ,  Singh J;
XX
DR   WPI; 1994-183419/22.
XX
PT   New peptide inhibitors of protein tyrosine kinase - contain D-Tyr
PT   or tetrafluoro-Tyr residue, for treating proliferative diseases,
PT   viral infection, inflammation etc.
XX
PS   Claim 7; Page 36; 44pp; English.
XX
CC   The sequence is that of a protein tyrosine kinase inhibitor contg. D-
CC   Tyr or tetrafluoro-Tyr residues.  The peptide can be used to inhibit
CC   TK-mediated signal transduction and is useful for controlling
CC   proliferative diseases, e.g. cancer, psoriasis and to treat viral,
CC   inflammatory, allergic and cardiovascular diseases.
CC   See also AAR55302-39.
XX
SQ   Sequence   9 AA;


  Query Match             45.9%;  Score 28;  DB 15;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 6.4e+05;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LATGSARQ 9
           || |||| :
Db       1 laegsare 8


RESULT   3
AAG95029
ID   AAG95029 standard; Peptide; 10 AA.
XX
AC   AAG95029;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 1223.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04776.
XX
PR   13-DEC-1999;   99GB-0029464.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-408419/43.
XX
PT   A set of peptide ligands consisting of specific complementary peptides
PT   to proteins encoded by genes of the human genome, useful in an assay
PT   for screening and identifying of one or more novel peptides which are
PT   drug candidates or pro-drugs -
XX
PS   Example 4; Page 215; 646pp; English.
XX
CC   The invention relates to a set of complementary peptide ligands
CC   generated from the human genome. The complementary peptides
CC   interact with their relevant target proteins encoded in the human
CC   genome. They can be used as reagents in drug discovery and as lead
CC   ligands to facilitate drug design and development. The present
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;


  Query Match             45.9%;  Score 28;  DB 22;  Length 10;
  Best Local Similarity   55.6%;  Pred. No. 79;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSARQ 9
           |  ||| |:
Db       1 cartgsgrr 9


RESULT   4
AAG95031
ID   AAG95031 standard; Peptide; 10 AA.
XX
AC   AAG95031;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 1225.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
```

PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 215; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 5
AAG95037
ID AAG95037 standard; Peptide; 10 AA.
XX
AC AAG95037;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1231.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 216; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 6
AAG95041
ID AAG95041 standard; Peptide; 10 AA.
XX
AC AAG95041;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1235.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 217; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 7
AAG95045

ID AAG95045 standard; Peptide; 10 AA.
XX
AC AAG95045;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1239.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 217; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 8
AAG95091
ID AAG95091 standard; Peptide; 10 AA.
XX
AC AAG95091;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1285.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 224; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 2 cartgsgrr 10

RESULT 9
AAG95093
ID AAG95093 standard; Peptide; 10 AA.
XX
AC AAG95093;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1287.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 224; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present

```
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;


  Query Match             45.9%;  Score 28;  DB 22;  Length 10;
  Best Local Similarity   55.6%;  Pred. No. 79;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CLATGSARQ 9
          |  ||| |:
Db      2 cartgsgrr 10


RESULT  10
AAG95095
ID   AAG95095 standard; Peptide; 10 AA.
XX
AC   AAG95095;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 1289.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04776.
XX
PR   13-DEC-1999;   99GB-0029464.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-408419/43.
XX
PT   A set of peptide ligands consisting of specific complementary peptides
PT   to proteins encoded by genes of the human genome, useful in an assay
PT   for screening and identifying of one or more novel peptides which are
PT   drug candidates or pro-drugs -
XX
PS   Example 4; Page 225; 646pp; English.
XX
CC   The invention relates to a set of complementary peptide ligands
CC   generated from the human genome. The complementary peptides
CC   interact with their relevant target proteins encoded in the human
CC   genome. They can be used as reagents in drug discovery and as lead
CC   ligands to facilitate drug design and development. The present
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;


  Query Match             45.9%;  Score 28;  DB 22;  Length 10;
  Best Local Similarity   55.6%;  Pred. No. 79;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CLATGSARQ 9
          |  ||| |:
Db      2 cartgsgrr 10


RESULT  11
AAG95097
ID   AAG95097 standard; Peptide; 10 AA.
XX
AC   AAG95097;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 1291.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04776.
XX
PR   13-DEC-1999;   99GB-0029464.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-408419/43.
XX
PT   A set of peptide ligands consisting of specific complementary peptides
PT   to proteins encoded by genes of the human genome, useful in an assay
PT   for screening and identifying of one or more novel peptides which are
PT   drug candidates or pro-drugs -
XX
PS   Example 4; Page 225; 646pp; English.
XX
CC   The invention relates to a set of complementary peptide ligands
CC   generated from the human genome. The complementary peptides
CC   interact with their relevant target proteins encoded in the human
CC   genome. They can be used as reagents in drug discovery and as lead
CC   ligands to facilitate drug design and development. The present
CC   sequence is a complementary peptide provided in the specification.
XX
SQ   Sequence   10 AA;


  Query Match             45.9%;  Score 28;  DB 22;  Length 10;
  Best Local Similarity   55.6%;  Pred. No. 79;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CLATGSARQ 9
          |  ||| |:
Db      1 cartgsgrr 9


RESULT  12
AAG95099
ID   AAG95099 standard; Peptide; 10 AA.
XX
AC   AAG95099;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Human complementary peptide, SEQ ID NO: 1293.
XX
KW   Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS   Homo sapiens.
XX
PN   WO200142277-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB04776.
XX
PR   13-DEC-1999;   99GB-0029464.
XX
PA   (PROT-) PROTEOM LTD.
XX
```

PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 225; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 13
AAG95101
ID AAG95101 standard; Peptide; 10 AA.
XX
AC AAG95101;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1295.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 226; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 14
AAG95115
ID AAG95115 standard; Peptide; 10 AA.
XX
AC AAG95115;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1309.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 228; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9
| ||| |:
Db 1 cartgsgrr 9

RESULT 15
AAG95141
ID AAG95141 standard; Peptide; 10 AA.
XX
AC AAG95141;
XX
DT 18-SEP-2001 (first entry)
XX

DE Human complementary peptide, SEQ ID NO: 1335.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs -
XX
PS Example 4; Page 232; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 CLATGSARQ 9
          |  ||| |:
Db      1 cartgsgrr 9
```

Search completed: June 6, 2002, 06:27:11
Job time: 204 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:25:32 ; Search time 12.84 Seconds
(without alignments)
22.828 Million cell updates/sec

Title: US-08-881-509-8
Perfect score: 61
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 83765

Minimum DB seq length: 0
Maximum DB seq length: 12

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	28	45.9	12	6	5424218-10	Patent No. 5424218
2	27	44.3	8	2	US-08-751-767A-80	Sequence 80, Appl
3	26	42.6	12	2	US-08-950-866-8	Sequence 8, Appli
4	25	41.0	9	1	US-08-049-794-27	Sequence 27, Appl
5	25	41.0	9	1	US-08-496-847-27	Sequence 27, Appl
6	25	41.0	9	2	US-08-742-774-27	Sequence 27, Appl
7	25	41.0	9	2	US-08-675-354-27	Sequence 27, Appl
8	25	41.0	9	2	US-08-965-918-27	Sequence 27, Appl
9	25	41.0	9	2	US-09-138-439-27	Sequence 27, Appl
10	25	41.0	9	3	US-08-613-400A-27	Sequence 27, Appl
11	25	41.0	9	3	US-09-298-017-27	Sequence 27, Appl
12	25	41.0	9	4	US-09-392-979A-27	Sequence 27, Appl
13	25	41.0	10	1	US-08-214-650-33	Sequence 33, Appl
14	24	39.3	7	6	5378464-26	Patent No. 5378464
15	24	39.3	10	3	US-08-159-339A-720	Sequence 720, App
16	24	39.3	12	1	US-07-834-902-6	Sequence 6, Appli
17	24	39.3	12	1	US-08-018-994-6	Sequence 6, Appli
18	24	39.3	12	1	US-08-294-675A-6	Sequence 6, Appli
19	24	39.3	12	1	US-08-397-633A-95	Sequence 95, Appl
20	24	39.3	12	2	US-08-950-866-3	Sequence 3, Appli
21	23	37.7	9	2	US-08-340-283-134	Sequence 134, App
22	22	36.1	9	1	US-08-445-745-136	Sequence 136, App
23	22	36.1	9	6	5217869-105	Patent No. 5217869
24	22	36.1	10	1	US-08-250-789A-107	Sequence 107, App
25	22	36.1	10	1	US-08-346-333-57	Sequence 57, Appl
26	22	36.1	10	4	US-09-230-548-6	Sequence 6, Appli
27	22	36.1	10	4	US-09-020-065A-26	Sequence 26, Appl
28	22	36.1	10	5	PCT-US91-07506-57	Sequence 57, Appl
29	22	36.1	12	1	US-08-548-540-151	Sequence 151, App
30	22	36.1	12	1	US-08-419-903A-6	Sequence 6, Appli
31	22	36.1	12	2	US-08-292-968-2	Sequence 2, Appli
32	22	36.1	12	2	US-08-467-974-2	Sequence 2, Appli
33	22	36.1	12	2	US-08-467-536-2	Sequence 2, Appli
34	22	36.1	12	3	US-08-467-976-2	Sequence 2, Appli
35	22	36.1	12	4	US-09-082-514-2	Sequence 2, Appli
36	22	36.1	12	5	PCT-US96-09809-151	Sequence 151, App
37	21	34.4	7	4	US-09-333-729A-18	Sequence 18, Appl
38	21	34.4	8	2	US-08-669-284B-32	Sequence 32, Appl
39	21	34.4	9	2	US-08-340-283-30	Sequence 30, Appl
40	21	34.4	9	2	US-08-340-283-104	Sequence 104, App
41	21	34.4	9	2	US-08-340-283-127	Sequence 127, App
42	21	34.4	9	4	US-09-258-754-333	Sequence 333, App
43	21	34.4	9	4	US-09-042-107-333	Sequence 333, App
44	21	34.4	9	4	US-08-584-008A-5	Sequence 5, Appli
45	21	34.4	9	4	US-09-492-543-33	Sequence 33, Appl

ALIGNMENTS

```
RESULT   1
5424218-10
;Patent No. 5424218
;     APPLICANT: MILJANICH, GEORGE P.;BITNER, ROBERT S.;BOWERSOX,
;. STEPHEN S.;FOX, JAMES A.;VALENTINO, KAREN L.;YAMASHIRO,DONALD H.
;     TITLE OF INVENTION: SCREENING METHOD FOR NEUROPROTECTIVE COMPOUNDS
;     NUMBER OF SEQUENCES: 21
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/147,714
;       FILING DATE: 04-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 855,269
;       FILING DATE: 23-MAR-1992
;       APPLICATION NUMBER: 561,766
;       FILING DATE: 02-AUG-1990
;       APPLICATION NUMBER: 440,094
;       FILING DATE:. 22-NOV-1989
;SEQ ID NO:10:
;       LENGTH: 12
5424218-10


  Query Match             45.9%;  Score 28;  DB 6;  Length 12;
  Best Local Similarity   62.5%;  Pred. No. 19;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSAR 8
           |  ||| |
Db       1 CCCTGSCR 8


RESULT   2
US-08-751-767A-80
; Sequence 80, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;     APPLICANT:  ANDERSON, ROBERT J.
;     APPLICANT:  GRANT, HUGH
;     APPLICANT:  MACDONALD, IAN D.
;     TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;     NUMBER OF SEQUENCES:  80
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NIXON & VANDERHYE P.C.
;       STREET:  1100 NORTH GLEBE ROAD
;       CITY:  ARLINGTON
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22201
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/751,767A
;       FILING DATE:  08-NOV-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SADOFF, B.J.
;       REGISTRATION NUMBER:  36,663
;       REFERENCE/DOCKET NUMBER:  117-221
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  80:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-751-767A-80


  Query Match             44.3%;  Score 27;  DB 2;  Length 8;
  Best Local Similarity   62.5%;  Pred. No. 1.7e+05;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CLATGSAR 8
           | |||  |
Db       1 CAATGRVR 8


RESULT   3
US-08-950-866-8
; Sequence 8, Application US/08950866
; Patent No. 5840296
;  GENERAL INFORMATION:
;    APPLICANT:  Raines, Ronald T
;    TITLE OF INVENTION:  Engineered Cytotoxic Ribonuclease A
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/950,866
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.94592
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-950-866-8


  Query Match             42.6%;  Score 26;  DB 2;  Length 12;
  Best Local Similarity   62.5%;  Pred. No. 47;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CLATGSAR 8
           |  |||:|
Db       1 CRDTGSSR 8


RESULT   4
US-08-049-794-27
; Sequence 27, Application US/08049794
; Patent No. 5587454
;  GENERAL INFORMATION:
;    APPLICANT:  JUSTICE, ALAN
;    APPLICANT:  SINGH, TEJINDER
;    APPLICANT:  GOHIL, KISHOR C
;    APPLICANT:  VALENTINO, KAREN L
;    APPLICANT:  MILJANICH, GEORGE P
;    TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;    TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Peter Dehlinger
;      STREET:  350 Cambridge Avenue, Suite 300
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,794
;      FILING DATE:  19930415
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/814,759
;      FILING DATE:  30-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A.
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  5865-0009.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
;      INDIVIDUAL ISOLATE:  32
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-049-794-27


  Query Match             41.0%;  Score 25;  DB 1;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7
```

```
RESULT   5
US-08-496-847-27
; Sequence 27, Application US/08496847
; Patent No. 5795864
;  GENERAL INFORMATION:
;     APPLICANT:  Amstutz, Gary A.
;     APPLICANT:  Bowersox, Stephen S.
;     APPLICANT:  Gohil, Kishorchandra
;     APPLICANT:  Adriaenssens, Peter I.
;     APPLICANT:  Kristipati, Ramasharma
;     TITLE OF INVENTION:  METHODS AND
;     TITLE OF INVENTION:  FORMULATIONS FOR PREVENTING PROGRESSION OF NEUROPATHIC PAIN
;     NUMBER OF SEQUENCES:  36
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dehlinger & Associates
;       STREET:  350 Cambridge Avenue, Suite 250
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  94306-1546
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/496,847
;       FILING DATE:  27-JUN-1995
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stratford, Carol A
;       REGISTRATION NUMBER:  34,444
;       REFERENCE/DOCKET NUMBER:  5865-0009.31
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-324-0880
;       TELEFAX:  650-324-0960
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  4
;       OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-496-847-27


  Query Match             41.0%;  Score 25;  DB 1;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT   6
US-08-742-774-27
; Sequence 27, Application US/08742774
; Patent No. 5824645
;  GENERAL INFORMATION:
;     APPLICANT:  JUSTICE, ALAN
;     APPLICANT:  SINGH, TEJINDER
;     APPLICANT:  GOHIL, KISHOR C
;     APPLICANT:  VALENTINO, KAREN L
;     APPLICANT:  MILJANICH, GEORGE P
;     TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;     TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;     NUMBER OF SEQUENCES:  34
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Peter Dehlinger
;       STREET:  350 Cambridge Avenue, Suite 300
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94306
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/742,774
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/675,354
;       FILING DATE:  03-JUL-1996
;       APPLICATION NUMBER:  US/08/049,794
;       FILING DATE:  1993-APR-15
;       APPLICATION NUMBER:  US 07/814,759
;       FILING DATE:  30-DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stratford, Carol A.
;       REGISTRATION NUMBER:  34,444
;       REFERENCE/DOCKET NUMBER:  5865-0009.30
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 324-0880
;       TELEFAX:  (415) 324-0960
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
;       INDIVIDUAL ISOLATE:  32
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  4
;       OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-742-774-27


  Query Match             41.0%;  Score 25;  DB 2;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT   7
US-08-675-354-27
; Sequence 27, Application US/08675354
; Patent No. 5859186
;  GENERAL INFORMATION:
;     APPLICANT:  JUSTICE, ALAN
;     APPLICANT:  SINGH, TEJINDER
;     APPLICANT:  GOHIL, KISHOR C
;     APPLICANT:  VALENTINO, KAREN L
;     APPLICANT:  MILJANICH, GEORGE P
;     TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;     TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;     NUMBER OF SEQUENCES:  34
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Peter Dehlinger
```

```
;      STREET:  350 Cambridge Avenue, Suite 300
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/675,354
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,794
;      FILING DATE:  1993-APR-15
;      APPLICATION NUMBER:  US 07/814,759
;      FILING DATE:  30-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A.
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  5865-0009.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
;      INDIVIDUAL ISOLATE:  32
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-675-354-27


  Query Match             41.0%;  Score 25;  DB 2;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT   8
US-08-965-918-27
; Sequence 27, Application US/08965918
; Patent No. 5891849
;  GENERAL INFORMATION:
;    APPLICANT:  Amstutz, Gary A.
;    APPLICANT:  Bowersox, Stephen S.
;    APPLICANT:  Gohil, Kishorchandra
;    APPLICANT:  Adriaenssens, Peter I.
;    APPLICANT:  Kristipati, Ramasharma
;    TITLE OF INVENTION:  METHODS AND FORMULATIONS FOR PREVENTING
;    TITLE OF INVENTION:  PROGRESSION OF NEUROPATHIC PAIN
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94306-1546
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/965,918
;      FILING DATE:  07-NOV-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mohr, Judy M.
;      REGISTRATION NUMBER:  38,563
;      REFERENCE/DOCKET NUMBER:  5865-0009.34
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-965-918-27


  Query Match             41.0%;  Score 25;  DB 2;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT   9
US-09-138-439-27
; Sequence 27, Application US/09138439
; Patent No. 5994305
;  GENERAL INFORMATION:
;    APPLICANT:  JUSTICE, ALAN
;    APPLICANT:  SINGH, TEJINDER
;    APPLICANT:  GOHIL, KISHOR C
;    APPLICANT:  VALENTINO, KAREN L
;    APPLICANT:  MILJANICH, GEORGE P
;    TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;    TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Peter Dehlinger
;      STREET:  350 Cambridge Avenue, Suite 300
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/138,439
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,794
;      FILING DATE:  1993-04-15
;      APPLICATION NUMBER:  US 07/814,759
```

```
;       FILING DATE:  30-DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stratford, Carol A.
;       REGISTRATION NUMBER:  34,444
;       REFERENCE/DOCKET NUMBER:  5865-0009.30
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 324-0880
;       TELEFAX:  (415) 324-0960
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
;       INDIVIDUAL ISOLATE:  32
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  4
;       OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-09-138-439-27


  Query Match             41.0%;  Score 25;  DB 2;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT  10
US-08-613-400A-27
; Sequence 27, Application US/08613400A
; Patent No. 6054429
;  GENERAL INFORMATION:
;    APPLICANT:  Bowersox, S. Scott
;    APPLICANT:  Gadbois, Theresa
;    APPLICANT:  Pettus, Mark, R.
;    APPLICANT:  Luther, Robert, R.
;    TITLE OF INVENTION:  IMPROVED EPIDURAL
;    TITLE OF INVENTION:  METHOD OF PRODUCING ANALGESIA
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94306-1546
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/613,400A
;      FILING DATE:  08-MAR-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  5865-0019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT,
;      INDIVIDUAL ISOLATE:  page 33, line29 to page 34, line4
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-08-613-400A-27


  Query Match             41.0%;  Score 25;  DB 3;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT  11
US-09-298-017-27
; Sequence 27, Application US/09298017
; Patent No. 6087091
;  GENERAL INFORMATION:
;    APPLICANT:  JUSTICE, ALAN
;    APPLICANT:  SINGH, TEJINDER
;    APPLICANT:  GOHIL, KISHOR C
;    APPLICANT:  VALENTINO, KAREN L
;    APPLICANT:  MILJANICH, GEORGE P
;    TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;    TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Peter Dehlinger
;      STREET:  350 Cambridge Avenue, Suite 300
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/298,017
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,794
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A.
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  5865-0009.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
```

```
;      INDIVIDUAL ISOLATE:  32
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-09-298-017-27


  Query Match             41.0%;  Score 25;  DB 3;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT  12
US-09-392-979A-27
; Sequence 27, Application US/09392979A
; Patent No. 6136786
;  GENERAL INFORMATION:
;    APPLICANT:  JUSTICE, ALAN
;    APPLICANT:  SINGH, TEJINDER
;    APPLICANT:  GOHIL, KISHOR C
;    APPLICANT:  VALENTINO, KAREN L
;    APPLICANT:  MILJANICH, GEORGE P
;    TITLE OF INVENTION:  METHODS OF PRODUCING ANALGESIA AND
;    TITLE OF INVENTION:  ENHANCING OPIATE ANALGESIA
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Peter Dehlinger
;      STREET:  350 Cambridge Avenue, Suite 300
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/392,979A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,794
;      FILING DATE:  1993-04-15
;      APPLICATION NUMBER:  US 07/814,759
;      FILING DATE:  30-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A.
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  5865-0009.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
;      INDIVIDUAL ISOLATE:  32
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  4
;      OTHER INFORMATION:  /note= "where X is hydroxyproline"
US-09-392-979A-27


  Query Match             41.0%;  Score 25;  DB 4;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.7e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CLATGSA 7
           ||: ||:
Db       1 CLSXGSS 7


RESULT  13
US-08-214-650-33
; Sequence 33, Application US/08214650
; Patent No. 5709995
;  GENERAL INFORMATION:
;    APPLICANT:  Chisari, Francis V.
;    APPLICANT:  Cerny, Andreas
;    TITLE OF INVENTION:  PEPTIDES FOR INDUCING CYTOTOXIC T
;    TITLE OF INVENTION:  LYMPHOCYTE RESPONSES TO HEPATITIS C VIRUS
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leydig, Voit & Mayer
;      STREET:  Two Prudential Plaza, Suite 4900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,650
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silvert, Donald J.
;      REGISTRATION NUMBER:  37552
;      REFERENCE/DOCKET NUMBER:  61230
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 616-5600
;      TELEFAX:  (312) 616-5700
;      TELEX:  25-3533
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-214-650-33


  Query Match             41.0%;  Score 25;  DB 1;  Length 10;
  Best Local Similarity   80.0%;  Pred. No. 60;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CLATG 5
           ||:||
Db       1 CLSTG 5


RESULT  14
5378464-26
;Patent No. 5378464
;    APPLICANT: MCEVER, RODGER P.
;    TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
;BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
;    NUMBER OF SEQUENCES: 32
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/320,408
```

; FILING DATE: 08-MAR-1989
;SEQ ID NO:26:
; LENGTH: 7
5378464-26

Query Match 39.3%; Score 24; DB 6; Length 7;
Best Local Similarity 80.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLATG 5
||||:|
Db 1 CLASG 5

RESULT 15
US-08-159-339A-720
; Sequence 720, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 720:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-720

Query Match 39.3%; Score 24; DB 3; Length 10;
Best Local Similarity 57.1%; Pred. No. 95;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TGSARQL 10
||: ||:
Db 4 TGAVRQI 10

Search completed: June 6, 2002, 06:27:31
Job time: 119 sec